DEC 27 1930

Index to Volume XXI (1930) in This Number

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

HAROLD RUGG, *Editor*

VOL. XXI, No. 9 — DECEMBER, 1930 — $5.00 per year

CONTENTS

COMPARATIVE RETENTION OF TYPEWRITING AND OF SUBSTITUTION WITH ANALOGOUS MATERIAL. *Frank N. Freeman and Ethel M. Abernethy* 639

THE CHARACTERISTICS OF SPELLING ERRORS. *James E. Mendenhall* 648

ON THE ANALYSIS OF CAUSATION. *Jack W. Dunlap and Edward E. Cureton* 657

CIRCUMSTANTIALITY AS A FACTOR IN GUESSING ON TRUE-FALSE EXAMINATIONS. *Ina Hill Brinkmeier and Noel Keys* . . . 681

STUDIES IN HANDEDNESS: II. TESTING BIMANUAL HANDEDNESS. *R. H. Ojemann* 695

STORY COMPLETION TESTS. *Grace H. Kent and E. Frances Wells* 703

SEX DIFFERENCES IN ALGEBRAIC ABILITY. *Glenn R. Pease* . . 712

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 715

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

ESTABLISHED 1910

Editor: HAROLD RUGG

Board of Associate Editors

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — E. L. THORNDIKE
FRANK N. FREEMAN — V. A. C. HENMON — L. M. TERMAN — GOODWIN B. WATSON

Assistant to the Editor: FRANCES M. FOSTER

The Journal is published monthly, except June to August, by Warwick and York, Inc., Baltimore, Md., and York, Pa.

The price is $5.00 a year, payable in advance; foreign postage 30 cents extra. Single current issues are 70 cents each, and subscriptions for less than a year are charged at single issue rate for each number ordered. Back volumes are $6.00 each; single back numbers, 90 cents each, except, when more than three years old, the price is $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence concerning editorial matters should be addressed to Harold Rugg, Editor, 425 West 123rd Street, New York City. Subscriptions should be sent to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*

TEACHERS' HEADQUARTERS

European Plan

Entirely Fire-Proof

Centrally Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

The Psychology of Learning
1928 Revision — BY WILLIAM H. PYLE
$2.20 + 12¢ postage
Warwick & York, Inc., *Baltimore, Maryland*

The Acid Test

What more exacting test can there be of dictionary leadership than the judgment of the educational world?

Every State that has adopted an unabridged dictionary for use in its schools has chosen exclusively

WEBSTER'S NEW INTERNATIONAL DICTIONARY

The Colleges voted overwhelmingly in favor of Webster as the standard of pronunciation, in answer to questions submitted by the Chicago Woman's Club.

Nearly 100% of all schoolbooks are based on the New International for pronunciation, spelling, compounding, and division of words.

The New International has been universally accepted by those best fitted to judge.

It is in accord with the best modern practice.

Use it in your own schools.

Write for our free booklets of interesting ways to teach the use of the dictionary.

GET THE BEST

G. & C. MERRIAM COMPANY
SPRINGFIELD, MASS.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXI December, 1930 Number 9

COMPARATIVE RETENTION OF TYPEWRITING AND OF SUBSTITUTION WITH ANALOGOUS MATERIAL

FRANK N. FREEMAN AND ETHEL M. ABERNETHY

PROBLEM

It is commonly said that motor habits deteriorate less with the lapse of time than do associations between words or other symbols. Swift and others have cited examples of the retention of such habits as ball tossing and typewriting over long periods of time with small loss. Such retention has been contrasted with the rapid loss of associations between nonsense syllables and the somewhat less rapid loss in memory of sense material.

The question has naturally been raised whether this difference in retention is due to the fact that one habit is motor and the other ideational, or at least that the overt movements are more prominent in the one than in the other activity, or whether the difference is due rather to the structure of the forms of learning which are compared. For example, the motor learning might consist of a small number of the same operations repeated over and over again, while the ideational learning consisted of a larger number of operations, each one of which was repeated less often. Thorndike refers to such a structural difference when he suggests that the greater permanence of motor learning may be due to a greater amount of overlearning.

The present experiment was planned with the purpose of setting up a comparison in retention between two forms of learning which should be as similar as possible in structure and should differ chiefly in the fact that the one involved overt movements and the other consisted primarily of associations of the ideational type. Except for this difference the content of the two forms of learning was to be the same. The first form, the motor learning, consisted in learning to typewrite a short paragraph with the keys covered with blank caps and a drawing of the keyboard placed in view of the learner. The

corresponding ideational learning consisted in translating the same paragraph into digits, following a key containing the letters of the alphabet and corresponding numbers placed before the learner.

Previous Studies

After the experiment was begun, the report of the experiment by McGeoch and Melton appeared.[1] These authors give a detailed review of the literature, and point out that no previous experiment made a direct comparison between motor learning and ideational learning. In their own experiment McGeoch and Melton compare the permanence of the ability to run three mazes with the permanence of the memory of three series of nonsense syllables. The interval between the first learning and the relearning was one week. On the basis of their results these authors challenge the common belief that motor learning possesses superior permanence. When retention was measured in terms of saving in trials the nonsense syllable lists were better retained than were the mazes. When the saving in time and errors was compared there was no marked difference. Recall scores calculated in terms of errors showed greater permanence of the maze habits, but they are not regarded as a valid basis of comparison.

The experiment of McGeoch and Melton makes a more direct comparison between overt motor learning and the formation of internal associations than had previously been set up. The significance of their experiment for our problem, however, is limited in two respects. In the first place, while the general conditions surrounding the two types of learning and the method of calculating the permanence of effect are similar, the two forms of learning differ radically in their structure. It is not possible to say, therefore, whether the presence of the overt movement in the one type of learning does or does not influence its permanence. In the second place, the interval between the learning and the relearning—one week—is too short to indicate what differences may exist throughout longer intervals.

Procedure

Since the two forms of learning were so closely similar, it was necessary to have different subjects learn the two tasks. Fifty sub-

[1] McGeoch, John A., and Melton, Arthur W.: The Comparative Retention Values of Maze Habits and of Nonsense Syllables. *Journal of Experimental Psychology*, Oct., 1929.

jects in all completed the learning and the first relearning. Twenty-five learned the typewriting and twenty-five the substitution. The two groups were equated approximately on the basis of mental-test scores (Terman Group Test). All the subjects were young women, members of college classes in psychology. None was at the beginning of the experiment in the least familiar with the keyboard of the typewriter.

Forty-four of these subjects, twenty-two in each group, completed, without any apparent irregularity in the experimental procedure, the second relearning. Six were rejected because of rehearsal after the first relearning or illness at the time for the second relearning.

Both groups learned the same material, consisting of a short paragraph which contained every letter of the alphabet.

Each individual of both groups worked in approximately fifty-minute practice periods a day until the task was completely learned. Each trial was timed in seconds by means of a stop watch.

The criterion of completion of the learning was the same for both groups—namely, two successive correct repetitions of the paragraph without the use of the diagram.

The detailed method of presentation to the typewriting group was as follows: The keys of the typewriter were covered with blank caps. A large diagram was kept on the wall in front of the subject until one errorless copy had been made. After this first correct copy the subject practised without the diagram. At the beginning of the first learning period the subjects were given the necessary information concerning the mechanism of the typewriter and the conventional position of the fingers upon the keyboard. All practice was upon the paragraph to be learned and every trial in practice was timed.

The presentation to the substitution group was made as follows: This group translated the paragraph into number symbols. On a small diagram, a copy of the large one used for the typewriting group, the letters were numbered according to their position on the keyboard. Numbers were also assigned for punctuation marks, the shift key, and the spacer, so that for every movement made in typing there was a particular number symbol. After going through the paragraph once silently, the subject recalled what she could, speaking the numbers aloud. The silent translation with the diagram and the recall aloud alternated until the subject could go through two successive correct recalls. At the beginning of subsequent learning periods a recall preceded the silent repetition.

The first relearning came two weeks after the completion of the original learning; the second, eight weeks after the first relearning. The procedure of relearning was the same as that of learning. In both the typewriting and the substitution the relearning was preceded by recall without the diagram. The criterion was the same as for the first learning—namely, two successive correct repetitions without the diagram.

The following were counted as errors in the case of both groups: Omissions, insertions, transportations, and substitutions. The same experimenter conducted the entire experiment. With each member of the typewriting group the experimenter was assisted by a teacher of typewriting, Miss Thelma Stone. The attempt was made to keep all conditions as constant as possible. Throughout the experiment the same room was used (for a particular subject), the same position in the room, and approximately the same time of day. Except for two subjects, all the experimental work was done in the same room and in the late afternoon. Subjects were warned at the close of each period to avoid rehearsing, and were questioned at the beginning of the following period. Whenever any condition was known to be irregular, the subject's record was discarded. Sixty subjects began the experiment, ten being eliminated.

It is clear that the situation and the learning process in the case of the two groups of learners were very similar. Both learned to translate the same verbal text. The number of items translated was therefore the same, and the perceptual and ideational content connected with the text was the same. Each letter, punctuation mark, and space of the text became associated with a definite reaction. In the one case an overt movement was a part of this reaction and in the other case it was not. Of course the overt movement was not all there was to the one reaction, and internal or incipient movements were not excluded from the other. But the overt movement constitutes a prominent distinguishing mark, and the reactions were very similar except for this distinction.

Results

The comparisons of the permanence of learning in the two groups are presented in terms of saving scores. The saving scores are calculated for the three measures, number of errors, number of repetitions, and time.

The mean scores and standard deviations in learning and first relearning of the two groups are presented in Table I. It will be seen that the number of errors made by the two groups in the original learning is approximately equal, but that typewriting required considerably more trials and longer time than did substitution.

The saving scores on the first relearning, at the end of two weeks, are given in Table II. The percentage of saving is remarkably similar in all three measures, number of errors, number of repetitions, and time. Thus far the results seem to agree with those of McGeoch and Melton.

Since three of each group dropped out before the second relearning, leaving twenty-two instead of twenty-five in each group, the mean scores of learning and first relearning, as well as of the second relearning, are presented for these groups. They are shown in Table III.

The saving scores for these groups are shown in Table IV. They do not differ significantly from those of the entire groups.

The saving scores of the second relearning, eight weeks after the first relearning, tell a different story. They are calculated in two ways, first taking as a base the means of the original learning, and, second, taking as a base the sum of the means of the original learning and the first relearning. The saving scores, calculated by these two methods, are given in Tables V and VI.

TABLE I.—THE AVERAGE SCORES IN LEARNING AND FIRST RELEARNING IN TYPEWRITING AND IN SUBSTITUTION. TWENTY-FIVE SUBJECTS IN EACH GROUP

Measure	Group	Errors		Repetitions		Time in seconds	
		Mean	SD	Mean	SD	Mean	SD
Learning	Substitution	158.80	123.74	17.24	6.10	3560.68	1105.47
	Typewriting	154.48	127.15	25.36	9.80	6358.52	1535.81
First relearning	Substitution	15.28	12.67	5.64	1.62	790.20	292.17
	Typewriting	17.52	11.05	8.24	3.27	1473.92	555.24

TABLE II.—SAVING SCORES BASED ON MEANS FROM TABLE I

Group	Errors	Repetitions	Time
Substitution	90.38	67.29	77.81
Typewriting	88.66	67.51	76.82

TABLE III.—THE AVERAGE SCORES IN LEARNING, IN FIRST RELEARNING, SECOND RELEARNING AND THE SUM OF THE SCORES IN LEARNING AND FIRST RELEARNING IN TYPEWRITING AND SUBSTITUTION. TWENTY-TWO SUBJECTS IN EACH GROUP

Measure	Group	Errors		Repetitions		Time in seconds	
		Mean	SD	Mean	SD	Mean	SD
Learning	Substitution	165.68	127.41	17.50	6.45	3596.59	1159.61
	Typewriting	136.36	84.41	25.68	10.06	6385.18	1503.97
First relearning	Substitution	14.95	13.32	5.59	1.72	792.36	311.06
	Typewriting	16.45	11.05	8.27	3.24	1477.36	559.28
Second relearning	Substitution	38.36	20.03	8.36	2.04	1228.00	364.81
	Typewriting	12.00	5.96	6.45	1.62	1242.59	124.26
Total learning and the first relearning	Substitution	180.63	129.44	23.09	7.24	4388.95	1311.74
	Typewriting	152.81	89.88	33.95	11.18	7862.54	1722.13

TABLE IV.—SAVING SCORES, THE FIRST RELEARNING IN RELATION TO LEARNING, BASED ON THE MEANS OF TABLE III

Group	Errors	Repetitions	Time
Substitution	90.98	68.06	77.97
Typewriting	87.94	67.80	76.86

TABLE V.—SAVING SCORES, THE SECOND RELEARNING IN RELATION TO LEARNING, BASED ON THE MEANS OF TABLE III

Group	Errors	Repetitions	Time
Substitution	76.85	52.23	65.86
Typewriting	91.20	74.88	80.54

TABLE VI.—SAVING SCORES, THE SECOND RELEARNING IN RELATION TO LEARNING PLUS THE FIRST RELEARNING, BASED ON THE MEANS OF TABLE III

Group	Errors	Repetitions	Time
Substitution	78.76	63.79	72.02
Typewriting	92.15	81.00	84.20

TABLE VII.—DIFFERENCES BETWEEN THE MEANS, FIRST RELEARNING AND SECOND RELEARNING, BASED ON THE MEANS OF TABLE III

	Substitution		Typewriting	
	D	PE of D	D	PE of D
Errors	+ 23.41	3.45	− 4.45	1.80
Repetitions	+ 2.77	.38	− 1.82	.51
Time in seconds	+435.64	68.94	−234.77	82.39

In view of the close similarity between the amount of retention after two weeks, the contrast after ten weeks is startling. In substitution, the loss which began during the first two weeks continued during the succeeding eight weeks. In typewriting, however, there was no further loss. On the contrary, there was some gain, for the saving scores on the second relearning were higher than on the first relearning. This gain is brought out in another way by a comparison of the mean scores of the first and second relearning, shown in Table VII.[1] The errors, repetitions, and time in substitution have all increased, whereas in typewriting they have all decreased. The difference is unmistakable. It seems to confirm without question the traditional belief that, for longer periods of time, learning which involves overt movements is more permanently retained than is learning which does not involve overt movements. The difference is so great that deterioration of the overt habit seems actually to be halted after two weeks.

It may be objected that the greater permanence of the typewriting is due to the fact that it required more trials and more time to learn this process. It may be thought that this constitutes overlearning. There was no overlearning in the typewriting, however, in comparison with substitution. The same number of analogous operations had to be learned in the two cases, and they were learned to the same degree.

The typewriting group did, to be sure, require more time, but this excess in time is allowed for by calculating the saving in terms of percentage. Furthermore, whatever effect may have resulted from this greater expenditure of time did not appear on the first relearning, after two weeks. It is the contrast between this equality in retention

[1] These differences in means are confirmed by differences in the number of individuals whose scores increased or decreased between the first and second learning.

during the first period and the marked inequality in retention during the second period which is important.

Again, it is noteworthy that the scores on the typewriting at the second relearning are actually higher than they were eight weeks before. The first relearning undoubtedly affected these scores; but the significant fact is that it did not similarly affect the scores in substitution.

Summary

This experiment was conducted to compare the permanence of "motor" learning with that of "ideational" learning. Earlier studies failed to settle the question because, first, the two types of learning were not directly compared under the same conditions, and, second, the cases of "motor" and "ideational" learning which were compared differed widely in structure as well as in the relative emphasis on overt movements. The recent experiment by McGeoch and Melton overcame the first difficulty but not the second.

In the present experiment the permanence of the effect of learning to typewrite a given text, using a diagram of the keyboard with the keys themselves covered, was compared with that of learning to substitute digits for letters with the same text and using a similar diagram with a digit added to accompany each letter. The same number of associations was thus learned in each of the two tasks. Each was learned to the point of two successive correct repetitions. There was therefore no more overlearning in the one case than in the other.

Each task was relearned after two weeks and again after an additional eight weeks. Twenty-five young women underwent the learning and first relearning of each task and twenty-two underwent the second relearning of each.

The typewriting required more time and a large number of trials to learn than did the substitution, but the percentage of saving in errors, time, and trials on the first relearning, after two weeks, was the same, within the limits of statistical error. For this period of time there was no difference in the permanence of the two types of learning. The similarity in the saving scores after the two-weeks interval is a check on the comparability of the two tasks.

After eight weeks more, or ten weeks from original learning, on the other hand, the amount which had been retained was very different. When the second relearning was compared with the original learning plus the first relearning, the percentage of saving in typewriting

exceeded that in substitution by 13.4 in errors, 17.2 in trials, and 12.2 in time. In substitution there was a marked loss in errors, trials, and time between the first and second relearning, whereas in typewriting there was a gain in each of these measures, varying from over twice the probable error to over three times the probable error.

The conclusion is that, for the types of learning studied, the existence of overt movements as part of a learning process makes the effect of learning more permanent over a ten-weeks period of time than when overt movements are not present, although it does not appear to affect the permanence over the shorter period of two weeks.

THE CHARACTERISTICS OF SPELLING ERRORS[1]

JAMES E. MENDENHALL

The Lincoln School of Teachers College, Columbia University

During the last three decades remarkable progress has been achieved in the field of English spelling. The content of spelling textbooks has been altered to include words that are commonly used by children and adults. Classroom methods have increasingly emphasized the learning of words in a natural context setting. Planned drills for initial learning and for relearning have been established as effective learning devices. Practice has been centered upon the words that are misspelled, rather than wastefully upon words already known. Systematic devices of self-drill and self-checking have promoted other economies of learning.

Despite this increase in knowledge and the consequent reforms, we are not completely informed about the exact nature of spelling errors and of the best manner to overcome them. In the present article an attempt has been made to contribute additional information to our knowledge of the nature of spelling errors. This information may have some important implications for the problem of spelling teaching.

In 1919 Horn[2] expressed a point of view relative to the character of misspellings for a particular word: "Recent investigations of the misspellings of pupils have shown that for most words there is no one error which is characteristic of a sufficiently large percentage of misspellings to warrant calling attention to it. Relatively few words have a type of error which includes as much as fifty per cent of the misspellings of pupils attempting those words. For most words, the percentage of the most frequent type of error to the total misspellings is very small." Since 1919 Horn has not seriously modified this statement.

---

[1] Some of the material for this article has been selected from "An Analysis of Spelling Errors," by James E. Mendenhall. The Lincoln School of Teachers College, Bureau of Publications, Teachers College, Columbia University, New York, 1930.

[2] Horn, Ernest: "Principles of Method in Teaching Spelling as Derived from Scientific Investigation." Eighteenth Yearbook of the National Society for the Study of Education, Part II, Public School Publishing Company, Bloomington, Ill., 1919, p. 67.

To the point the writer analyzed letter by letter a sample of one hundred eighty-four words administered to one hundred pupils in Grade III, and a sample of ninety-two words given to one hundred pupils in Grade IV. In each sample the word list was one grade level above that of the pupils—that is, a fourth grade list of words was presented to Grade III. This device produced for the analysis a larger number of mistakes than would have been normally secured.

Most studies of spelling errors have considered only the most frequent error for a word. In the present study every error for a word was analyzed. Conclusions about the commonness of a particular error will depend upon which type of analysis is employed. For example, if only the most frequent error for the word "trouble" is analyzed, the omission of the letter "u" characterizes forty per cent of the total errors for the word. If all of the errors are analyzed, the omission of the letter "u" characterizes sixty per cent of the total errors. Table I displays the means of the per cents for the analysis of the most frequent error (partial analysis) and for the analysis of all of the errors (the complete analysis). In Grade III the per cent is raised from twenty-three to forty-eight; in Grade IV, from twenty-six to forty-five. These are significant increases. (These per cents would have been signally larger had the word lists been of the same grade level as the pupils.) Although the reader will decide for himself, the writer believes that these per cents are sufficiently large to warrant special attention in spelling procedures.

Another matter of possible interest to students of spelling is displayed in Fig. 1. This figure shows the number of errors in successive letter positions of words of various lengths. For example, for the word "trouble" there were no errors in the "t" or first position, five errors in the "r" or second position, eight errors in the "o" or third position, etc. These data for "trouble" were assembled with similar data for other words of seven letters. For each letter position means were computed. From these means Fig. 1 was constructed. This figure exhibits a number of important facts. In words of any length, there is a marked increase in the number of errors from the first to the last letter position. In an average four-letter word the ratio of the number of errors in the first position to the number in the last position is one to three; in an average nine-letter word, the ratio is one to thirty. In words of any length the position of letters with the greatest number of errors is at the center or directly to the right of center of the word. The number of letters in a word does not seem to influence the difficulty

of the initial letters. In fact pupils can spell the initial four letters of a nine-letter word as easily as the initial four letters of a five-letter word.

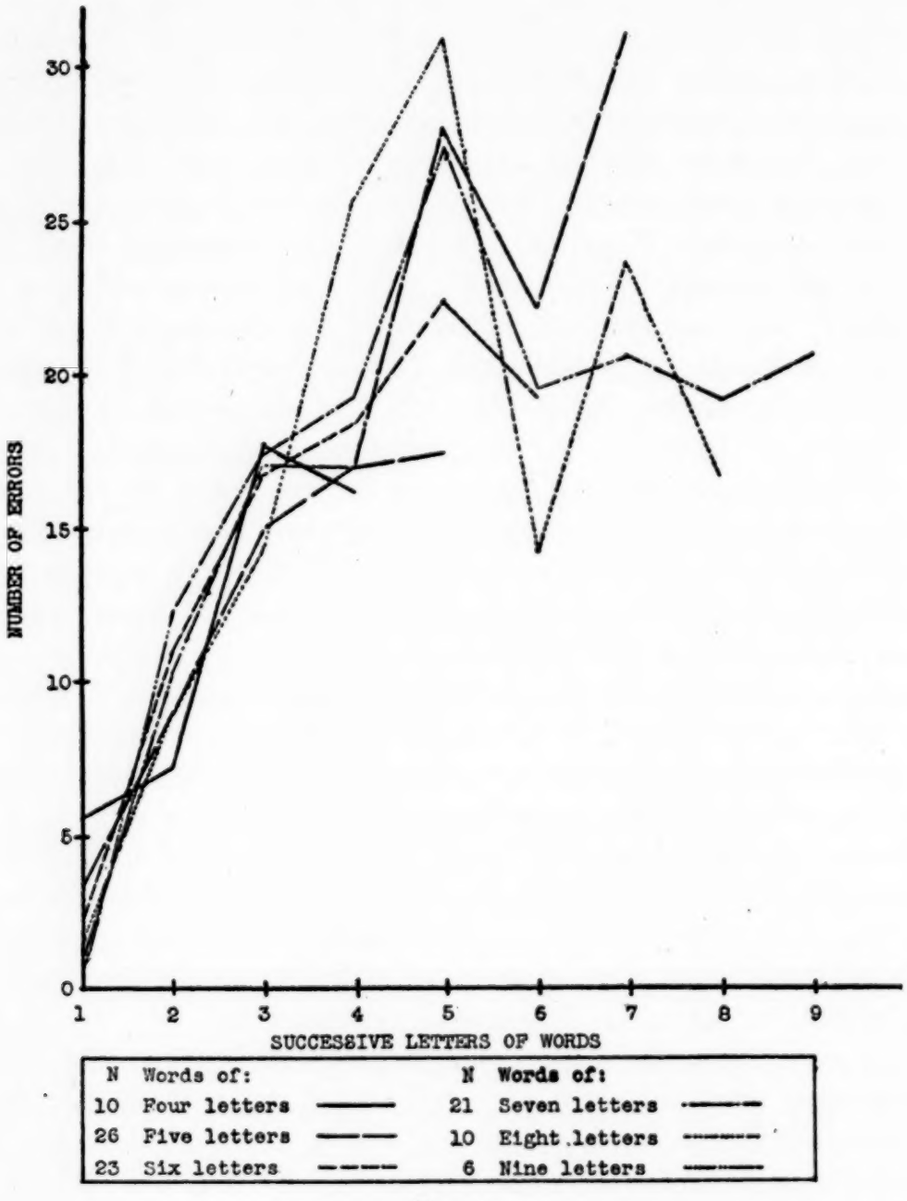


FIG. 1.

In the sample from Grade III the number of errors in one letter position was correlated with the number of errors in every other position and with the difficulty of the word. For six-letter words, these data appear in Table II. The original correlations appear in Roman type; the partial correlations (difficulty constant) appear in italics.

TABLE I.—MEANS AND STANDARD DEVIATIONS OF PER CENT OF TOTAL ERRORS FOR WORDS ACCOUNTED FOR BY MOST FREQUENT ERROR, MOST FREQUENT ERROR DETERMINED BY (*A*) PARTIAL ANALYSIS OF ERRORS AND (*B*) COMPLETE ANALYSIS OF ALL ERRORS

	Per cent total errors, partial analysis	Per cent total errors, complete analysis	Difference
Grade III (List IV) $N = 184$:			
Mean	23.04	48.29	25.25
SD	11.54	17.69	1.18
Grade IV (List V) $N = 92$:			
Mean	25.95	44.97	19.02
SD	29.91	17.86	3.07
	Grade III	Grade IV	
Correlation coefficient (per cent partial and per cent complete)	.466	.324	
SD	.06	.09	

TABLE II.—CORRELATIONS AMONG NUMBER OF ERRORS IN SUCCESSIVE LETTERS OF FORTY-THREE SIX-LETTER WORDS, GRADE III (LIST IV)[1]

Successive letters	2	3	4	5	6	Difficulty
1	.350	.220	.301	.255	.141	.413
	.241	*−.044*	*.034*	*.025*	*.105*	
2		.490	.304	.210	.136	.560
		.227	*−.120*	*−.162*	*−.243*	
3			.513	.060	−.027	.609
			.175	*−.443*	*−.528*	
4				.319	.298	.674
				−.110	*−.102*	
5					.559	.572
					.364	
6						.536

[1] Original correlations in Roman type. First order partial correlations (difficulty constant) in italic type.

From the original correlations it appears that the association between the number of errors in adjacent positions is distinctly higher than those between the numbers in non-adjacent positions. When a letter in any position is in error, there is a decided tendency to err in the immediately succeeding letter. The correlation between the number of errors in any letter position and difficulty is significantly higher than the correlation between the numbers in any two letter positions, adjacent or non-adjacent.

From the partial correlations (difficulty constant) practically all correlations are materially reduced in size. The tendency to misspell adjacent letters is greater in the first part of a word than in the last part of the word, except in the case of the last two letters. If the second or third letter is misspelled, later letters are generally correct.

Altogether the writer administered 2300 words to one hundred public-school children, Grades I–VI. For every word, the most frequent error was determined. For the word "trouble," this error was "troble," the omission of the letter "u" being the principal source of error. Counting the letter "u" in the word "trouble" as one unit, the total frequency of error for this letter was compiled. A similar procedure was followed in the case of other letters. The second column of Table III portrays the per cents of the grand total based upon an assembly of the letters most frequently in error.

In a sample of 10,000 letters (from any 2000 words, averaging five letters in length), the letter "e" would occur 1300 times, or thirteen per cent of the total times. (The third column of Table III shows the per cents for the letter "e" and for other letters.) Since the letter "e" accounts for thirteen per cent of 10,000 letters, we would normally expect it to occasion thirteen per cent of the errors in words. This expectation is supported by the high rank-order coefficient of correlation of .853 between error and occurrence. However, such a correlation obscures the true nature of the case. The letter "e," occurring thirteen per cent of the time, is associated with twenty-three per cent of all the errors, a difference of ten per cent. From the fourth column of Table III it can be observed that the letters "e," "a," "i," and "u" are most frequently associated with error, and that the letters "t," "n," "h," and "r" are least frequently accompaniments of error. It is interesting to note that the four most difficult letters are vowels, and that the four least difficult letters are consonants.

There are other aspects of words that are slightly but not significantly associated with the difficulty of words. The brevity of this

article precludes the incorporation of tables to display the supporting data. However, the conclusions from these tables may be of some interest to the reader. The length of words is only slightly associated with spelling difficulty; difficulty is a function of particular letters (singly or combined) rather than of the gross number of letters. The

TABLE III.—FREQUENCY OF ERROR AND OF OCCURRENCE OF VARIOUS LETTERS OF THE ALPHABET

Letter	Error in 1,857 letters, per cent	Occurrence in 10,000 letters,[1] per cent	Difference (occurrence-error), per cent
a	13.2	7.78	− 5.42
b	1.0	1.41	.41
c	4.6	2.96	− 1.64
d	3.4	4.02	.62
e	22.9	12.77	−10.13
f	.6	1.97	1.37
g	1.0	1.74	.74
h	1.9	5.95	4.05
i	11.0	6.67	− 4.33
j	.3	.57	.27
k	1.3	.74	− .56
l	3.5	3.72	.22
m	1.0	2.88	1.88
n	2.3	6.86	4.56
o	7.4	8.07	.67
p	1.5	2.23	.73
q	.1	.08	− .02
r	3.3	6.51	3.21
s	5.1	6.22	1.12
t	2.7	8.55	5.85
u	6.9	3.08	− 3.82
v	.2	1.12	.92
w	2.2	1.76	− .44
x	.2	.27	.07
y	2.5	1.96	− .54
z	.0	.17	.17
	100.1	100.06	
rho (error and occurrence)			.853 ± .04

[1] The per cents were computed from the letters and their frequencies which appeared in "Secret Messages, How to Read and How to Write Them," p. 23, by Paul B. Thomas. Alfred A. Knopf, New York, 1929.

number of vowels, the number of syllables, the position of the accent, the confusion of words—none of these exhibits a consistent or high association with word difficulty.

Let us turn from facts of general scientific import to facts of particular significance to the teaching of spelling. Hollingworth[1] has expressed an opinion that prevails in the field: "*English spelling must be acquired by the formation of a comparatively few fundamental bonds.*" Horn[2] justifies this position by citing the unphonetic character of English spelling. However, he introduces an important qualification: "It is probable that many children who have not been taught phonics have made their own comparisons between words and have in that way built up a sort of phonic system of their own." In spelling, therefore, would it not be the wise course to guide the development of these generalizations rather than to leave their formation to chance?

To further the growth of generalizations, two possible methods of grouping words present themselves—grouping by like auditory elements, and grouping by like visual elements. The writer's study showed that fifty-one per cent of all errors for words are reasonably phonetic renditions. These errors that are phonetically correct would not be reduced or eliminated by an auditory method of grouping. Studies by Gates[3] emphasize the importance of visual perception in spelling, and imply that visual grouping should prove most effective in learning.

The writer made a study of the errors for three hundred fifty words from Grade IV that may have some bearing upon word grouping in spelling instruction. The words were classified by common types of error. Practically all of the words for a grade fell within the following categories: (1) The *addition* or *omission* (*a*) of the second of a double letter, (*b*) of a final letter, (*c*) of silent letters, (*d*) of key consonants, and (*e*) of central syllables; and (2) the *substitution* (*a*) of other words for the correct ones, and (*b*) of one vowel for another in the second

---

[1] Hollingworth, Leta S.: "The Psychology of Special Disability in Spelling." New York: Bureau of Publications, Teachers College, Columbia University, Contributions to Education, No. 88, 1918, p. 65.

[2] Horn, Ernest: The Child's Early Experience with the Letter A. *Journal of Educational Psychology*, Vol. XXI, March, 1929, pp. 161–162.

[3] Gates, Arthur I.: "The Psychology of Reading and Spelling." New York: Bureau of Publications, Teachers College, Columbia University, Contributions to Education, 1922, p. 86; and Gates, Arthur I.: A Study of the Rôle of Visual Perception, Intelligence, and Certain Associative Processes in Reading and Spelling. *Journal of Educational Psychology*, Vol. XVII, October, 1926, p. 444.

syllable of words. This classification suggested the use of common errors as a basis for grouping. For example, words like "address," "banner," "slippers" (and others from Category 1*a*) might be taught together. These words have a common visual feature (the double letter) and a common error (the omission of one member of a double letter). Likewise, another group of words might be formed to include "chain," "gain," "grain," "maid," and "pain." These words exhibit a common visual element ("ai") and a common type of error (the omission of the letter "i"). The words, "ankle," "bridle," "couple," and "gentle" have a common error (the transposition of the letters "l" and "e"). Another group would incorporate "apron," "cabin," "comfort," "habit," "palace," and "postal." These words commonly have errors in the second syllable, in which it is difficult to ascertain the vowel from the sound of the syllable. In this last case the system of grouping rests upon the empirical study of error rather than upon common visual features of words.

In essence the writer suggests a visual method of instruction that is based upon the common errors for words. The visual method of grouping alone reveals so many possibilities that the organizer is at a loss to select the optimal one. However, if commonality of error is made the basis, the problem of grouping is greatly simplified.

Summary

The conclusions and implications from the study can be stated succinctly as follows:

1. The most frequent error for a word characterizes fifty per cent of the misspellings for that word. This per cent is high enough to warrant special consideration.

2. In general, the hard spot of a word is either at the center or immediately at the right of center. The initial letters of a short word are no more difficult to spell than the initial letters of a long word.

3. The number of errors in any one letter position are associated with the number in any other position less closely than with the difficulty of the word.

4. In general, the letters occurring most frequently are most often in error. Moreover, some letters "e," "a," "i," and "u" are associated with error much more frequently than their occurrence warrants.

5. Difficulty is associated with particular letters (singly or combined) more closely than with general aspects of words, such as length,

the number of vowels, the number of syllables, and the position of the accent.

6. The writer suggests the use of a method of grouping based upon two features of words and their errors: (*a*) Common visual elements and (*b*) common types of errors. The efficacy of this method, of course, awaits further experimental investigation.

ON THE ANALYSIS OF CAUSATION

JACK W. DUNLAP AND EDWARD E. CURETON

Territorial Normal and Training School, Honolulu, Hawaii

I. Introduction

Much has been written about the indiscriminate use of statistics in the analysis of the causes underlying observed phenomena in various fields. Many of these problems of causation are not, however, insoluble. Statistical procedures have been developed by many workers during the last several years which when properly applied are capable of throwing much light on them, but which when improperly applied lead to serious errors in interpretations and conclusions. These contributions have appeared in the special journals of many different fields, and at irregular intervals. The purpose of this paper is to bring together some of these techniques, add certain others, point out a few of the relationships between them, and show their applications to various problems. The illustrative problems considered hereafter will be taken from the major fields of interest of the writers, namely education and psychology.

II. The Principle of the Analysis of Variance

Most problems of causation resolve themselves fundamentally into a consideration of the proportions of the variability in an effect attributable to the different causes under consideration. These problems can best be attacked by the methods of the analysis of variance and by other techniques which depend for their significance upon their relationships with these methods.

Let x_0 be a score (measured from its mean, $\bar{x}_0$) in the "dependent" variable (criterion or effect).

Let $x_1, x_2, \ldots x_n$ be scores (measured from their respective means, $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_n$) in the "independent"* variables (causes).

Consider the simplest case of causation, where

$$x_0 = x_1 + x_2.$$

Then

$$\sigma_0^2 = \sigma^2_{(1+2)} = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 r_{12}.$$

* An "independent" variable is to be thought of as one whose observed values are used to predict values of a "dependent" variable. Mathematical independence is not implied here.

If x_1 and x_2 are uncorrelated, $r_{12} = 0$, and

$$\sigma_0^2 = \sigma_1^2 + \sigma_2^2 \qquad (1)$$

This equation (1) is the basic formula for the analysis of variance. The variance of any variable may be defined as the mean square deviation about the arithmetic mean, or as the square of the standard deviation. Equation (1) may then be expressed as follows:

If every score in a dependent (mathematically) variable be composed of the sum of the scores in two independent (mathematically) variables, then the variance of the dependent variable equals the sum of the variances of the independent variables. This rule may readily be extended to cover the case of n mutually independent variables. Furthermore, if any of these be composed of m other mutually independent variables (on which observations are available), the same rule may be applied to effect an analysis of these remote causes. The ratio of the variance of one of the independent variables to that of the criterion gives a measure of the proportion of the criterion (considered as an effect) due to that variable (considered as a cause). The sum of all such ratios will of course equal one.

The fact that one variable may be the sum of several others does not mean necessarily that it is to be considered as an effect. In fact, no mathematical relationship in itself can ever indicate the *direction* of causation; this must be inferred from the nature of the experimental conditions or on other *a priori* grounds.

III. Correlation

If we think of one variable as an effect and of another as a partial cause, we may separate the variance of the effect into two portions, one of which is uncorrelated with that of the partial cause while the other is perfectly correlated with it. In particular, the variance of the effect may be separated into one part representing the variance of the observations about the regression line (and hence uncorrelated with the partial cause), and another part representing the variance of points on the regression line about the mean of the effect (hence, determined completely by the variance of the partial cause).

Let x_0 be the effect.

$_1x_0$ that part of the effect determined by the partial cause.

$x_{0.1}$ that part of the effect independent of the partial cause.

x_1 the partial cause, all being measured from their respective means.

$$x_0 = {}_1x_0 + x_{0.1}$$

$$r_{01} = r_{x_0({}_1x_0)}$$

$$r_{01} = r_{_1x_0(_1x_0 + x_{0.1})} = \frac{\Sigma_1 x_0^2 + \Sigma_1 x_0(x_{0.1})}{N_1\sigma_0\sigma_0}$$

but $r_{_1x_0(x_{0.1})} = 0$ by definition, so

$$r_{01} = \frac{{}_1\sigma_0}{\sigma_0} \tag{2}$$

Stated verbally: *The correlation between an effect and an assumed partial cause is equal to the ratio of that portion of the standard deviation of the effect which is determined by the partial cause to the total standard deviation of the effect.* The former standard deviation may be found from the equation

$${}_1\sigma_0 = b_{01}\sigma_1 \tag{3}$$

where b_{01} is the slope of the regression line of x_0 on x_1.

The residual standard deviation, $\sigma_{0.1}$, is the standard error of estimate.

$$\sigma_0^2 = {}_1\sigma_0^2 + \sigma^2_{0.1},$$

and from (2),

$$r^2_{01} = \frac{{}_1\sigma_0^2}{\sigma_0^2} = \frac{{}_1\sigma_0^2}{{}_1\sigma_0^2 + \sigma^2_{0.1}}. \tag{4}$$

This equation (4) expresses the relation between the correlation coefficient and the proportions of the variance of the criterion which are respectively determined by, and independent of, that of the "independent" variable. Obviously, if ${}_1\sigma_0 = 0$, $r_{01} = 0$, and if $\sigma_{0.1} = 0$, $r_{01} = 1$.

$$r^2_{01} = \frac{{}_1\sigma_0^2}{\sigma_0^2} = \frac{\sigma_0^2 - \sigma^2_{0.1}}{\sigma_0^2}$$

$$r_{01} = \sqrt{1 - \frac{\sigma^2_{0.1}}{\sigma_0^2}}. \tag{5}$$

$$\sigma^2_{0.1} = \sigma_0^2 - \sigma_0^2 r^2_{01}$$

$$\sigma_{0.1} = \sigma_0\sqrt{1 - r^2_{01}} = k_{01}\sigma_0 \tag{6}$$

Equation (6) is the ordinary formula for the standard error of estimate, and (5) is a formula relating the standard error of estimate and the standard deviation of the "dependent" variable to the correlation between the "independent" and "dependent" variables.

In the case of linear regression, ${}_1\sigma_0$ may be determined directly from (3), and the correlation calculated from the ordinary product-moment formula, but if the regression is non-linear, equation (3) breaks down, and must be replaced by an equation involving higher powers of the variables, the nature of the equation depending on the form of the regression line. In this case the product-moment formula

breaks down also, as it is based essentially on the assumption of linearity of regression. However, equation (5) is not affected by the form of the regression line; hence it may be used to calculate the correlation when the regression is non-linear. The interpretation in terms of the analysis of variance by equation (4) still holds. The fitting of curved regression lines and the computation of the standard error of estimate and the index of correlation for such lines have been treated by Mills.[1] He has pointed out that equation (5) describes the correlation between two variables no matter what the form of the regression line, the correlation coefficient representing the special case where this is a straight line fitted by the method of least squares, and the correlation ratio the special case where the line is arbitrarily made to pass through the means of the x_0-arrays. The writers have not seen the interpretation in terms of the analysis of variance—equation (4)—pointed out except for the two special cases of the correlation coefficient and ratio.

An important special case is that in which the partial cause contributes itself completely—*i.e.*, in which ${}_1x_0 = cx_1$, c being a constant. In this case,

$$r_{01} = \frac{c\sigma_1}{\sigma_0}. \tag{7}$$

$$r^2_{01} = \frac{c^2{\sigma_1}^2}{{\sigma_0}^2} = \frac{c^2{\sigma_1}^2}{c^2{\sigma_1}^2 + {\sigma^2}_{0\cdot 1}}. \tag{8}$$

and

$$k^2_{01} = 1 - \frac{c^2{\sigma_1}^2}{{\sigma_0}^2}.$$

If the units of measurement are assumed or made equal, c will equal unity. Under these conditions, r^2_{01} gives the proportion of ${\sigma_0}^2$ which is determined by ${\sigma_1}^2$ and k^2_{01} the proportion of ${\sigma_0}^2$ determined by residual factors.

Tryon[2] has pointed out that the above special case has in fact been assumed by many writers to be a general statement; but that the use of r^2 and k^2 as the ratios of the variance of the criterion determined respectively by the "independent" variable and residual factors is only correct when the "independent" variable is known or assumed to contribute itself completely. The following analysis is adapted from Tryon.

We have so far *assumed* that one variable was the effect and the other the partial cause. The correlation coefficient itself gives no clue as to which variable should be considered as causal, nor does it

determine entirely the extent to which the causal variable contributes. We have just considered the case where the latter contributes itself completely, and found that r^2 gives the proportion of its contribution. This relation holds whichever variable is assumed to be the causal one, and hence represents an important limiting case. The general case is the one in which the two variables under consideration are both affected by a common cause. Calling the variables correlated x_0 and x_1, and the causal factor x_2, we have,

$$x_0 = c_0 x_2 + x_{0.2} \qquad \sigma_0{}^2 = c_0{}^2\sigma_2{}^2 + \sigma^2{}_{0.2}$$
$$x_1 = c_1 x_2 + x_{1.2} \qquad \sigma_1{}^2 = c_1{}^2\sigma_2{}^2 + \sigma^2{}_{1.2}$$

Noting that x_2, $x_{0.2}$ and $x_{1.2}$ are uncorrelated,

$$r_{01} = \frac{c_0 c_1 \Sigma x_2{}^2}{N\sigma_0\sigma_1} = \frac{c_0 c_1 \sigma_2{}^2}{\sigma_0\sigma_1}. \tag{9}$$

Also

$$r_{02} = \frac{c_0\sigma_2}{\sigma_0}, \text{ and } r_{12} = \frac{c_1\sigma_2}{\sigma_1},$$

so that

$$r_{02}r_{12} = \frac{c_0 c_1 \sigma_2{}^2}{\sigma_0\sigma_1} = r_{01} \tag{10}$$

Formula (7) can be shown to be a special case of formula (9). If x_1 is determined completely by x_2,

$$x_1 = c_1 x_2$$
$${}_2x_0 = {}_1x_0$$
$$x_{0.2} = x_{0.1}$$

But from our definition in the general case,

$$x_0 = c_0 x_2 + x_{0.2},$$

whence

$$c_0 x_2 = {}_2x_0 = {}_1x_0.$$

Now from the first of our special conditions,

$$x_2 = \frac{x_1}{c_1},$$

so that

$${}_1x_0 = \frac{c_0}{c_1}x_1,$$

and

$$\frac{c_0}{c_1} = c$$

(of formula (7)), and

$$\sigma_2^2 = \frac{\sigma_1^2}{c_1^2}.$$

Then substituting the value of σ_2^2 in (9),

$$r_{01} = \frac{c_0 c_1 \sigma_1^2}{c_1^2 \sigma_0 \sigma_1} = \frac{\frac{c_0}{c_1}\sigma_1}{\sigma_0} = \frac{c\sigma_1}{\sigma_0},$$

which is identical with formula (7).

The other limiting special case is that in which the residual correlations in the two variables under consideration are equal. In this case $r_{02}^2 = r_{12}^2$, and from (10)

$$r_{01} = r^2_{02} = r^2_{12}. \qquad (11)$$

From (7) and (11) it is apparent that we can pass to the proposition that under these conditions r_{01} gives the proportion of σ_0^2 which is determined by σ_1^2, or *vice versa*. Thus it is seen that the proportion of the variance of one variable determined by any other may always be represented by a fraction whose value lies somewhere between that of the correlation coefficient and its square. Methods for evaluating this fraction have been discussed by Tryon. In applying these results to the estimation of the community of function between variables measured by fallible observations, it is necessary to use correlation coefficients corrected for attenuation.

A common fallacy in computing correlations is the use of quotients involving the chronological age in the denominator as data. Such studies, involving correlations between intelligence quotients, etc., are all likely to suffer from spurious index correlation caused by the common variation of the chronological age in the paired observations. It should be noted that correlating such quotients does not result in eliminating mutual correlation with age, except in certain special cases. This matter was discussed in an early paper by Pearson[3] and in a later one by Thomson and Pintner.[4]

IV. Partial Correlation

The use of partial correlation in the analysis of causation is beset with many pitfalls, as has been pointed out by Burks.[5, 6] One of these is the attempt to argue directly from partial correlation to causation. Partial coefficients do not indicate the direction of causation any more than do total coefficients, or any other statistical measures of con-

comitant variation. All that such measures provide is an estimate of the magnitude of the relationship; never of the direction of causation.

The partial correlation coefficient measures the correlation between that part of a variable which is uncorrelated with one or more others and that part of a second variable which is also uncorrelated with these others. It "eliminates" not only the effects of the variable or variables partialed out upon the variables whose residual correlation is to be determined, but also any effects of the correlated variables upon those "eliminated." Hence, in general, the partial correlation coefficient is likely in most cases to eliminate more than is desired. An important exception is the case of chronological age. Obviously, this must always operate as a "cause" when used to represent maturity, and there are many cases in which it is desired to estimate the relationship between two traits both of which exhibit progressive development with age. If the only observations available are from groups heterogeneous as to age, the partial correlation technique is appropriate.

There is one special difficulty encountered in dealing with correlations between a variable such as a test score or a physical measurement and chronological age. This arises from the fact that the regressions are significantly curvilinear if the age range becomes at all large (more than one or two years). If the only reason for computing such coefficients is for use in "partialing out" the chronological age, it is not necessary to fit curved regression lines and calculate indexes of correlation about them. Fisher[7] has pointed out for the case of the time series, which is essentially similar, that the effect of time (in this case age) may be eliminated by partialing out time, its square, its cube, etc., treating these as separate variates. For most practical purposes, it is unnecessary to go beyond the cubes, and in a great many cases the squares.

V. Semi-partial Correlation

There arise many problems in the study of causation where it is desirable to know the correlation between that part of one variable which is uncorrelated with one or more others and another variable or that part of it which is uncorrelated with one or more others. The "one or more others" may or may not be entirely or partly the same in both cases. If they are entirely the same we have the case of Yule's partial correlation.[8] If they are partly or entirely different, we shall designate the coefficient as a semi-partial coefficient. An example of the use of semi-partial correlation might be the following:

Given the intelligence scores of a number of children, the intelligence scores of their parents, and a rating of the home environment (such as would be given by the Sims Socio-Economic Score Card). It is desired to know whether the home environment affects the child's intelligence score more than is to be expected from a knowledge of the fact that the more intelligent parents are likely to provide the better homes. Now it is obvious that the parents' intelligence will affect the home environment, and that the latter will probably affect the child's intelligence score. We may assume that the effect of the child's intelligence on the home environment will be negligible. The problem then reduces itself to estimating the correlation between that part of the home environment score which is uncorrelated with the parents' intelligence score and the child's intelligence score. The ordinary partial correlation between home environment and child's intelligence with parents' intelligence "held constant" is not applicable because it represents the correlation between that part of home environment uncorrelated with parents' intelligence and *that part of child's intelligence uncorrelated with parents' intelligence.* The semi-partial correlation appropriate to the solution of this problem will be designated $r_{(1.3)2}$, the correlation between that part of variable 1 uncorrelated with 3, and all of 2.

The following analysis represents an attempt to derive the principal systems of semi-partial correlation, of which Yule's formula is one, and the above is the three-variable case of another.

Let

$$z_1 = \frac{x_1}{\sigma_1},\ z_2 = \frac{x_2}{\sigma_2}.$$

Then

$$\sigma_{z_1} = \sigma_{z_2} = \cdots = \sigma_{z_n} = 1 \tag{12}$$

$$r_{12} = \frac{\Sigma x_1 x_2}{N\sigma_1\sigma_2} = \frac{\Sigma(x_1/\sigma_1)(x_2/\sigma_2)}{N} = \frac{\Sigma z_1 z_2}{N} \tag{13}$$

Let

$$x_{1.2} = x_1 - r_{12}\frac{\sigma_1}{\sigma_2}x_2.$$

$x_{1.2}$ represents a deviation (in the x_1-dimension) from the regression line of x_1 on x_2 (assuming this regression to be linear). If we consider all the observations on any subject as coordinates $x_1, x_2, x_3, \ldots x_n$ of a point in an n-dimensional continuum, then $x_{1.2}, x_3, \ldots x_n$ will represent the coordinates in an $(n-1)$-dimensional continuum in which

the value given by the regression line of x_1 on x_2 replaces the mean of x_1 as the origin of measurement of $x_{1.2}$ in the x_1-dimension, and the x_2-dimension does not appear. Similarly $x_{1.23}$, x_4, . . . x_n, will represent the coordinates in an $(n-2)$-dimensional continuum in which the value given by the regression plane of x_1 on x_2 and x_3 is taken as the origin for $x_{1.23}$ in the x_1-dimension, and in general $x_{1.23\ldots m}$, x_{m+1}, x_{m+2}, . . . x_n (where $m < n$) will represent the coordinates in an $(n-m+1)$-dimensional continuum in which the value given by the $(m-1)$-dimensional regression hyperplane is taken as the origin for $x_{1.23\ldots m}$ in the x_1-dimension, and the dimensions numbered 2 to m inclusive do not appear.*

$$\sigma_{1.2} = \sqrt{\frac{\Sigma\left(x_1 - r_{12}\frac{\sigma_1}{\sigma_2}x_2\right)^2}{N}} = \sigma_1\sqrt{1-r^2_{12}} \tag{14}$$

Let

$$z_{1.2} = \frac{x_{1.2}}{\sigma_{1.2}} = \frac{x_1 - r_{12}\frac{\sigma_1}{\sigma_2}x_2}{\sigma_1\sqrt{1-r^2_{12}}} = \frac{z_1 - r_{12}z_2}{\sqrt{1-r^2_{12}}} \tag{15}\dagger$$

$$r_{(1.3)(2.3)} = \frac{\Sigma z_{1.3}z_{2.3}}{N} = \frac{\Sigma(z_1 - r_{13}z_3)(z_2 - r_{23}z_3)}{N\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

$$= \frac{\Sigma z_1z_2 - r_{13}\Sigma z_2z_3 - r_{23}\Sigma z_1z_3 + r_{13}r_{23}\Sigma z_3{}^2}{N\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

From (12) and (13),

$$r_{(1.3)(2.3)} = \frac{r_{12} - r_{13}r_{23} - r_{13}r_{23} + r_{13}r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

$$r_{(1.3)(2.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}} = r_{12.3} \tag{16}$$

This is Yule's formula for the case of three variables.

We shall in general put a z, with subscripts, for the ratio of an x with the same subscripts to its corresponding standard deviation.

---

* The writers are indebted to Dr. Harold Hotelling for assistance in this geometrical interpretation.

† This is not the same as the $z_{1.2}$ used by T. L. Kelley[9] (p. 284) in deriving the multiple regression equation. He takes as the unit of measurement the standard deviation of the criterion, so that $z_{1.2} = x_{1.2}/\sigma_1 = z_1 - r_{12}z_2$, while in the present analysis, the standard error of estimate is the unit of measurement used, the regression line becoming the origin instead of the mean.

Then,

$$z_{1.23} = \frac{x_{1.23}}{\sigma_{1.23}} = \frac{x_{1.3} - r_{12.3}\frac{\sigma_{1\,3}}{\sigma_{2.3}}x_{2.3}}{\sigma_{1.3}\sqrt{1 - r^2_{12.3}}}$$

$$z_{1.23} = \frac{z_{1.3} - r_{12.3}z_{2.3}}{\sqrt{1 - r^2_{12.3}}} \tag{17}$$

and it can be shown similarly that,

$$z_{1.234\ldots n} = \frac{z_{1.34\ldots n} - r_{12.34\ldots n}z_{2.34\ldots n}}{\sqrt{1 - r^2_{12.34\ldots n}}} \tag{18}$$

Total correlation coefficients will be hereafter designated as r_{12}, etc., as usual; partial coefficients $r_{12.34\ldots n}$ as suggested by Yule; and semi-partial coefficients $r_{(1.34)(2.56)}$, etc., the two terms in parentheses being the things correlated, and each representing that part of the variable before the dot which is uncorrelated with those after the dot. Thus the example here given should be read, "the correlation between that part of variable 1 uncorrelated with 3 and 4, and that part of variable 2 uncorrelated with 5 and 6." In this notation, Yule's partial coefficient would be written $r_{(1.34\ldots n)(2.34\ldots n)}$.

It may be noted that in the following derivations no attempt is made to indicate economical methods of computation. The formulas presented may indicate the method of building up semi-partials from others of the next lower order, but the selection of the particular coefficients of each order to use in order to perform the computations with the least possible number of arithmetical operations has not been investigated, nor has the matter of numerical checks been considered. The writers hope to deal with these matters in a later paper.

System A—Partial Correlation

$$r_{(1.34)(2.34)} = \frac{\Sigma z_{1.34}\, z_{2.34}}{N} = \frac{\Sigma(z_{1.4} - r_{13.4}\, z_{3.4})(z_{2.4} - r_{23.4}\, z_{3.4})}{N\sqrt{1 - r^2_{13.4}}\sqrt{1 - r^2_{23.4}}}$$

$$= \frac{\Sigma z_{1.4}\, z_{2.4} - r_{23.4}\Sigma z_{1.4}\, z_{3.4} - r_{13.4}\Sigma z_{3.4}\, z_{2.4} + r_{13.4}r_{23.4}\Sigma z^2_{3.4}}{N\sqrt{1 - r^2_{13.4}}\sqrt{1 - r^2_{23.4}}}$$

$$= \frac{r_{12.4} - r_{23.4}r_{13.4} - r_{13.4}r_{23.4} + r_{13.4}r_{23.4}}{\sqrt{1 - r^2_{13.4}}\sqrt{1 - r^2_{23.4}}}$$

$$r_{(1.34)(2.34)} = \frac{r_{12.4} - r_{13.4}r_{23.4}}{\sqrt{1 - r^2_{13.4}}\sqrt{1 - r^2_{23.4}}} = r_{12.34} \tag{19}$$

Yule's formula in the four variable case. It is not necessary to derive any other special cases of partial correlation, so we shall pass directly to the general case. Yule's $r_{12.34\ \ldots\ n}$ is in our notation

$$r_{(13.4\ \cdots\ n)(2.34\ \cdots\ n)} = \frac{\Sigma z_{1.34\ \cdots\ n} z_{2.34\ \cdots\ n}}{N}$$

$$= \frac{\Sigma(z_{1.45\ \cdots\ n} - r_{13.45\ \cdots\ n} z_{3.45\ \cdots\ n})(z_{2.45\ \cdots\ n} - r_{23.45\ \cdots\ n} z_{3.45\ \cdots\ n})}{N\sqrt{1 - r^2_{13.45\ \cdots\ n}}\sqrt{1 - r^2_{23.45\ \cdots\ n}}}$$

$$= (\Sigma z_{1.45\ \cdots\ n} z_{2.45\ \cdots\ n} - r_{23.45\ \cdots\ n} \Sigma z_{1.45\ \cdots\ n} z_{3.45\ \cdots\ n} - r_{13.45\ \cdots\ n}$$
$$\Sigma z_{3.45\ \cdots\ n} z_{2.45\ \cdots\ n} + r_{13.45\ \cdots\ n} r_{23.45\ \cdots\ n} \Sigma z^2_{3.45\ \cdots\ n}) \div$$
$$(N\sqrt{1 - r^2_{13.45\ \cdots\ n}}\sqrt{1 - r^2_{23.45\ \cdots\ n}})$$

$$= \frac{r_{12.45\ \cdots\ n} - r_{23.45\ \cdots\ n} r_{13.45\ \cdots\ n} - r_{13.45\ \cdots\ n} r_{23.45\ \cdots\ n} + r_{13.45\ \cdots\ n} r_{23.45\ \cdots\ n}}{\sqrt{1 - r^2_{13.45\ \cdots\ n}}\sqrt{1 - r^2_{23.45\ \cdots\ n}}}$$

$$r_{(1.34\ \cdots\ n)(2.34\ \cdots\ n)} = \frac{r_{12.45\ \cdots\ n} - r_{13.45\ \cdots\ n} r_{23.45\ \cdots\ n}}{\sqrt{1 - r^2_{13.45\ \cdots\ n}}\sqrt{1 - r^2_{23.45\ \cdots\ n}}} \qquad (20)$$

This is the general case. It is interesting to note here that the coefficients of each order are expressed in terms of those of the next lower order until finally we have the ordinary total coefficients. It should be noted further that in the following derivations the same general trend is found, except that coefficients are expressed in semi-partials as well as partials of the next lower order.

System B

The correlation between that part of variable 1 uncorrelated with 3, and all of variable 2 may be derived as follows:

$$r_{(1.3)2} = \frac{\Sigma z_{1.3} z_2}{N} = \frac{\Sigma(z_1 - r_{13} z_3) z_2}{N\sqrt{1 - r^2_{13}}} = \frac{\Sigma z_1 z_2 - r_{13} \Sigma z_2 z_3}{N\sqrt{1 - r^2_{13}}}$$

$$r_{(1.3)2} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r^2_{13}}} \qquad (21)^*$$

For the case of four variables,

$$r_{(1.34)2} = \frac{\Sigma z_{(1.34)} z_2}{N} = \frac{\Sigma(z_{1.4} - r_{13.4}\, z_{3.4}) z_2}{N\sqrt{1 - r^2_{13.4}}}$$

$$= \frac{\Sigma z_{1.4} z_2 - r_{13.4} \Sigma z_{3.4} z_2}{N\sqrt{1 - r^2_{13.4}}}$$

$$r_{(1.34)2} = \frac{r_{(1.4)2} - r_{13.4} r_{(3.4)2}}{\sqrt{1 - r^2_{13.4}}} \qquad (22)$$

---

* This special case has been derived by Franzen[10] by a slightly different method from the one given here.

In the general case,

$$r_{(1.34\ \cdots\ n)2} = \frac{\Sigma z_{1.34\ \cdots\ n} z_2}{N} = \frac{\Sigma(z_{1.4\ \cdots\ n} - r_{13.4\ \cdots\ n} z_{3.4\ \cdots\ n}) z_2}{N\sqrt{1 - r^2_{13.4\ \cdots\ n}}}$$

$$= \frac{\Sigma z_{1.4\ \cdots\ n} z_2 - r_{13.4\ \cdots\ n} \Sigma z_{3.4\ \cdots\ n} z_2}{N\sqrt{1 - r^2_{13.4\ \cdots\ n}}}$$

$$r_{(1.34\ \cdots\ n)2} = \frac{r_{(1.4\ \cdots\ n)2} - r_{13.4\ \cdots\ n} r_{(3.4\ \cdots\ n)2}}{\sqrt{1 - r^2_{13.4\ \cdots\ n}}} \qquad (23)$$

This is the case where that part of variable 1 uncorrelated with n-2 others is correlated with all of variable 2.

System C

The correlation between that part of variable 1 which is uncorrelated with variables 3 and 4 and that part of variable 2 which is uncorrelated with 3 may be derived as follows:

$$r_{(1.34)(2.3)} = \frac{\Sigma z_{1.34} z_{2.3}}{N} = \frac{\Sigma(z_{1.4} - r_{13.4} z_{3.4})(z_2 - r_{23} z_3)}{N\sqrt{1 - r^2_{13.4}}\sqrt{1 - r^2_{23}}}$$

$$r_{(1.34)(2.3)} = \frac{r_{(1.4)2} - r_{23} r_{(1.4)3} - r_{13.4}\, r_{(3.4)2} + r_{23} r_{13.4}\, r_{(3.4)3}}{\sqrt{1 - r^2_{13.4}}\sqrt{1 - r^2_{23}}} \qquad (24)$$

Continuing to the case which is general for one variable,

$$r_{(1.34\ \cdots\ n)(2.3)} = \frac{\Sigma z_{1.34\ \cdots\ n} z_{2.3}}{N} = \frac{\Sigma(z_{1.4\ \cdots\ n} - r_{13.4\ \cdots\ n} z_{3.4\ \cdots\ n})(z_2 - r_{23} z_3)}{N\sqrt{1 - r^2_{13.4\ \cdots\ n}}\sqrt{1 - r^2_{23}}}$$

$$r_{(1.34\ \cdots\ n)(2.3)} =$$

$$\frac{r_{(1.4\ \cdots\ n)2} - r_{23} r_{(1.4\ \cdots\ n)3} - r_{13.4\ \cdots\ n} r_{(3.4\ \cdots\ n)2} + r_{23} r_{13.4\ \cdots\ n} r_{(3.4\ \cdots\ n)3}}{\sqrt{1 - r^2_{13.4\ \cdots\ n}}\sqrt{1 - r^2_{23}}} \qquad (25)$$

This is the case of the correlation between that part of variable 1 which is uncorrelated with 3, 4, etc. (to any number), and that part of 2 which is uncorrelated with 3.

In the general case, where $m < n$,

$$r_{(1.34\ \cdots\ n)(2.34\ \cdots\ m)} = \frac{\Sigma z_{1.34\ \cdots\ n} z_{2.34\ \cdots\ m}}{N} =$$

$$\frac{\Sigma(z_{1.4\ \cdots\ n} - r_{13.4\ \cdots\ n} z_{3.4\ \cdots\ n})(z_{2.4\ \cdots\ m} - r_{23.4\ \cdots\ m}\, z_{3.4\ \cdots\ m})}{N\sqrt{1 - r^2_{13.4\ \cdots\ n}}\sqrt{1 - r^2_{23.4\ \cdots\ m}}}$$

$$r_{(1.34\ \cdots\ n)(2.34\ \cdots\ m)} = (r_{(1.4\ \cdots\ n)(2.4\ \cdots\ m)} - r_{23.4\ \cdots\ m}\, r_{(1.4\ \cdots\ n)(3.4\ \cdots\ m)}$$

$$- r_{13.4\ \cdots\ n}\, r_{(3.4\ \cdots\ n)(2.4\ \cdots\ m)} + r_{13.4\ \cdots\ n}\, r_{23.4\ \cdots\ m}\, r_{(3.4\ \cdots\ n)(3.4\ \cdots\ m)})$$

$$\div (\sqrt{1 - r^2_{13.4\ \cdots\ n}}\sqrt{1 - r^2_{23.4\ \cdots\ m}}) \qquad (26)$$

This is the general case for the correlation between that part of variable 1 which is uncorrelated with 3, 4, etc. (to any number); and that part of 2 which is uncorrelated with some but not all of these variables. In evaluating the formulas of System C we use total correlations and semi-partials of Systems A, B, and D.

SYSTEM D

$$r_{(1.2)1} = \frac{\Sigma z_{1.2}\, z_1}{N} = \frac{\Sigma(z_1 - r_{12} z_2) z_1}{N\sqrt{1 - r^2_{12}}}$$

$$= \frac{\Sigma z_1^2 - r_{12}\, \Sigma z_1 z_2}{N\sqrt{1 - r^2_{12}}} = \frac{1 - r^2_{12}}{\sqrt{1 - r^2_{12}}}$$

$$r_{(1.2)1} = \sqrt{1 - r^2_{12}} \tag{27}$$

Formula (27) gives the correlation between that part of variable 1 which is uncorrelated with variable 2 and all of 1. This, as one would expect, is Kelley's k, the coefficient of alienation, or in terms of standard scores the standard error of estimate.

$$r_{(1.23\,\cdots\,n)1} = \frac{\Sigma z_{1.23\,\cdots\,n}\, z_1}{N} = \frac{\Sigma(z_{1.3\,\cdots\,n} - r_{12.3\,\cdots\,n} z_{2.3\,\cdots\,n}) z_1}{N\sqrt{1 - r^2_{12.3\,\cdots\,n}}}$$

$$r_{(1.23\,\cdots\,n)1} = \frac{r_{(1.34\,\cdots\,n)1} - r_{12.34\,\cdots\,n}\, r_{(2.34\,\cdots\,n)1}}{\sqrt{1 - r^2_{12.34\,\cdots\,n}}} \tag{28}$$

or

$$r_{(1.23\,\cdots\,n)1} = \sqrt{1 - r^2_{1(23\,\cdots\,n)}} \tag{29}$$

The coefficient in (29) is a multiple correlation coefficient. Formulas (28) and (29) give the correlation between that part of variable 1 which is uncorrelated with variables 2, 3, etc. (to any number), and all of 1. This special case is the general multiple alienation coefficient.

Where $m < n$,

$$r_{(1.2\,\cdots\,n)(1.2\,\cdots\,m)} = \frac{\Sigma z_{1.2\,\cdots\,n}\, z_{1.2\,\cdots\,m}}{N}$$

$$= \frac{\Sigma(z_{1.3\,\cdots\,n} - r_{12.3\,\cdots\,n} z_{2.3\,\cdots\,n})(z_{1.3\,\cdots\,m} - r_{12.3\,\cdots\,m} z_{2.3\,\cdots\,m})}{N\sqrt{1 - r^2_{12.3\,\cdots\,n}}\sqrt{1 - r^2_{12.3\,\cdots\,m}}}$$

$$r_{(1.2\,\cdots\,n)(1.2\,\cdots\,m)} = (r_{(1.3\,\cdots\,n)(1.3\,\cdots\,m)} - r_{12.3\,\cdots\,m}\, r_{(1.3\,\cdots\,n)(2.3\,\cdots\,m)}$$

$$- r_{12.3\,\cdots\,n}\, r_{(2.3\,\cdots\,n)(1.3.\,\cdots\,m)} + r_{12.3\,\cdots\,n}\, r_{12.3\,\cdots\,m}\, r_{(2.3\,\cdots\,n)(2.3\,\cdots\,m)}) \div (\sqrt{1 - r^2_{12.3\,\cdots\,n}}\sqrt{1 - r^2_{12.3\,\cdots\,m}}) \tag{30}$$

This is the general case of the correlation between that part of variable 1 which is uncorrelated with several others and that part of 1 which is uncorrelated with some but not all of these others.

System E

$$r_{(1.3)(2.4)} = \frac{\Sigma z_{1.3}\, z_{2.4}}{N} = \frac{\Sigma(z_1 - r_{13}\, z_3)\,(z_2 - r_{24}\, z_4)}{N\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{24}}}$$

$$r_{(1.3)(2.4)} = \frac{r_{12} - r_{14}\, r_{24} - r_{13}\, r_{23} + r_{13}\, r_{24}\, r_{34}}{\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{24}}} \qquad (31)^*$$

In the general case where n is odd and m is even, odd and even subscripts designating different variables,

$$r_{(1.357\,\cdots\,n)(2.468\,\cdots\,m)} = \frac{\Sigma z_{1.35\,\cdots\,n} z_{2.46\,\cdots\,m}}{N}$$

$$= \frac{\Sigma(z_{1.57\,\cdots\,n} - r_{13.57\,\cdots\,n} z_{3.57\,\cdots\,n})(z_{2.68\,\cdots\,m} - r_{24.68\,\cdots\,m}\, z_{4.68\,\cdots\,m})}{N\sqrt{1 - r^2_{13.57\,\cdots\,n}}\sqrt{1 - r^2_{24.68\,\cdots\,m}}}$$

$$r_{(1.357\,\cdots\,n)(2.468\,\cdots\,m)} = (r_{(1.57\,\cdots\,n)(2.68\,\cdots\,m)} - r_{24.68\,\cdots\,m} r_{(1.57\,\cdots\,n)(4.68\,\cdots\,m)} - r_{13.57\,\cdots\,n} r_{(3.57\,\cdots\,n)(2.68\,\cdots\,m)} + r_{13.57\,\cdots\,n} r_{24.68\,\cdots\,m} r_{(3.57\,\cdots\,n)(4.68\,\cdots\,m)}) \div (\sqrt{1 - r^2_{13.57\,\cdots\,n}}\sqrt{1 - r^2_{24.68\,\cdots\,m}}) \qquad (32)$$

System F

The correlation between that part of variable 1 uncorrelated with variable 2, and all of 2, may be derived as follows:

$$r_{(1.2)2} = \frac{\Sigma z_{1.2} z_2}{N} = \frac{\Sigma(z_1 - r_{12} z_2) z_2}{N\sqrt{1 - r^2_{12}}} = \frac{r_{12} - r_{12}}{\sqrt{1 - r^2_{12}}}$$

$$r_{(1.2)2} = 0 \qquad (33)$$

Note that variable 1 may be extended to cover the case of a battery of observations weighted according to the multiple regression equation, so that the correlation between a criterion (in this case variable 2) and that part of the battery uncorrelated with the criterion becomes

$$r_{([1345\,\cdots\,n].2)2} = 0 \qquad (34)$$

---

* This special case has been derived by Frazen[10] by a slightly different technique.

$$r_{(1.23\ \ldots\ n)2} = \frac{\Sigma z_{1.23\ \ldots\ n} z_2}{N}$$

$$= \frac{\Sigma(z_{1.3\ \ldots\ n} - r_{12.3\ \ldots\ n} z_{2.3\ \ldots\ n}) z_2}{N\sqrt{1 - r^2_{12.3\ \ldots\ n}}}$$

$$r_{(1.23\ \ldots\ n)2} = \frac{r_{(1.3\ \ldots\ n)2} - r_{12.3\ \ldots\ n} r_{(2.3\ \ldots\ n)2}}{\sqrt{1 - r^2_{12.3\ \ldots\ n}}} = 0 \quad (35)$$

The correlation between that part of variable 1 which is uncorrelated with variable 3 and that part of variable 3 which is uncorrelated with variable 2 is derived as follows:

$$r_{(1.3)(3.2)} = \frac{\Sigma z_{1.3} z_{3.2}}{N} = \frac{\Sigma(z_1 - r_{13} z_3)(z_3 - r_{23} z_2)}{N\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{23}}}$$

$$= \frac{r_{13} - r_{12} r_{23} - r_{13} + r_{13} r^2_{23}}{\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{23}}}$$

$$r_{(1.3)(3.2)} = \frac{r_{23}(r_{13} r_{23} - r_{12})}{\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{23}}}. \quad (36)$$

The presence of correlation in this case may be explained by the analysis of variance.

$$\sigma^2_{1.3} = \sigma^2_{1.(3.2)} + \sigma^2_{1.(23)}$$

This may be read: "That part of the variance of variable 1 which is independent of variable 3 may be thought of as being composed of two portions: that part of the variance of 1 which is independent of the part of the variance of 3 which is independent of 2, and that part of the variance of 1 which is independent of the part of the variance of 3 which is determined by 2." The dot, as usual, signifies independence of succeeding variables; the previous subscript denotes dependence.

Then from equation (2),

$$r_{(1.3)(3.2)} = \frac{\sigma_{1.(3.2)}}{\sigma_{1.3}} > 0, \text{ if } \sigma_{1.(3.2)} > 0,$$

i.e., if variable 3 is not completely determined by 2.

$$r_{(1.34\ \ldots\ n)(3.2)} = \frac{\Sigma z_{1.34\ \ldots\ n} z_{3.2}}{N}$$

$$= \frac{\Sigma(z_{1.4\ \ldots\ n} - r_{13.4\ \ldots\ n} z_{3.4\ \ldots\ n})(z_3 - r_{23} z_2)}{N\sqrt{1 - r^2_{13.4\ \ldots\ n}}\sqrt{1 - r^2_{23}}}$$

$$r_{(1.34\ \ldots\ n)(3.2)} = (r_{(1.4\ \ldots\ n)3} - r_{23} r_{(1.4\ \ldots\ n)2} - r_{13.4\ \ldots\ n} r_{(3.4\ \ldots\ n)3} + r_{23} r_{13.4\ \ldots\ n} r_{(3.4\ \ldots\ n)2}) \div (\sqrt{1 - r^2_{13.4\ \ldots\ n}}\sqrt{1 - r^2_{23}}) \quad (37)$$

In the general case, where n is odd and m is even,

$$r_{(1.35\ldots n)(3.24\ldots m)} = \frac{\Sigma z_{1.35\ldots n} z_{3.24\ldots m}}{N}$$

$$= \frac{\Sigma(z_{1.57\ldots n} - r_{13.57\ldots n} z_{3.57\ldots n})(z_{3.46\ldots m} - r_{32.46\ldots m} z_{2.46\ldots m})}{N\sqrt{1 - r^2_{13.57\ldots n}}\sqrt{1 - r^2_{32.46\ldots m}}}$$

$$r_{(1.35\ldots n)(3.24\ldots m)} = (r_{(1.57\ldots n)(3.46\ldots m)} - r_{32.46\ldots m} r_{(1.57\ldots n)(2.46\ldots m)} - r_{13.57\ldots n} r_{(3.57\ldots n)(3.46\ldots m)} + r_{13.57\ldots n} r_{32.46\ldots m} r_{(3.57\ldots n)(2.46\ldots m)}) \div (\sqrt{1 - r^2_{13.57\ldots n}}\sqrt{1 - r^2_{32.46\ldots m}}) \quad (38)$$

VI. Multiple Correlation and Path Coefficients

The essential problem of this section is to determine the ratio of the portion of the standard deviation of a criterion determined by one of the "independent" variables in a system to the total standard deviation of this criterion; or to determine the proportion of the total variance of the criterion (considered as an effect) ascribable to the "independent" variable (considered as a cause).

It has already been shown (Sec. III) that in the case of a criterion and one "independent" variable, the ratio of the part of the standard deviation of the criterion which is determined by its regression on the "independent" variable to its total standard deviation is equal to the coefficient of correlation between the criterion and the "independent" variable. The analogous coefficient, where there are other "independent" variables in the system, has been termed a path coefficient by Wright.[11] The proportion of the variance of the criterion which is determined by its regression on one of the "independent" variables has been designated by Wright as the coefficient of determination of the criterion by the given "independent" variable. The coefficient of determination is thus the square of the corresponding path coefficient, and, by the principle of the analysis of variance, the sum of all such coefficients must equal one, if the system includes all the causes of variation in the criterion. In any ordinary system the total variance of the criterion will not be determined completely by the "independent" variables. The ratio of the residual variance in the criterion to its total variance will equal the square of T. L. Kelley's[9] (p. 288) multiple alienation coefficient; and the ratio of the

sum of all the portions of the variance of the criterion determined by the "independent" variables of the system to its total variance will equal the square of the multiple correlation coefficient.

Let x_0 be a criterion score (measured from the mean) and x_1, x_2, . . . x_n corresponding scores in the "independent" variables.

If there is only one "independent" variable,

$$ {}_1x_0 = r_{01}\frac{\sigma_0}{\sigma_1}x_1 $$

(so that ${}_1x_0 + x_{0.1} = x_0$), the pre-subscript indicating the "determining" variable. We should read ${}_1x_0$ as "that part of x_0 determined by its regression on x_1." All regressions are assumed to be linear.

Similarly, it can be shown that

$$ \left.\begin{aligned} {}_{1.2}x_0 &= \frac{r_{0(1.2)}\sigma_0}{\sigma_{1.2}}x_1 \\ {}_{2.1}x_0 &= \frac{r_{0(2.1)}\sigma_0}{\sigma_{2.1}}x_2 \end{aligned}\right\} \tag{39} $$

and in general

$$ {}_{1.23\ \cdots\ n}x_0 = \frac{r_{0(1.23\ \cdots\ n)}\sigma_0}{\sigma_{1.23\ \cdots\ n}}x_1 \tag{40} $$

where

$$ \sigma_{1.23\ \cdots\ n} = \sigma_{1.23\ \cdots\ (n-1)}\sqrt{1 - r^2_{1n.23\ \cdots\ (n-1)}} $$

${}_{1.23\ \cdots\ n}x_0$ should be read, the part of x_0 determined by its regression on the part of x_1 which is independent of x_2, x_3, and other "independent" variables. Then

$$ {}_1\sigma_0 = r_{01}\frac{\sigma_0}{\sigma_1}\sigma_1 = r_{01}\sigma_0 \tag{41} $$

$$ {}_{1.2}\sigma_0 = \frac{r_{0(1.2)}\sigma_0\sigma_1}{\sigma_{1.2}} $$

$$ {}_{2.1}\sigma_0 = \frac{r_{0(2.1)}\sigma_0\sigma_2}{\sigma_{2.1}} $$

$$ {}_{1.23\ \cdots\ 1n}\sigma_0 = \frac{r_{0(1.23\ \cdots\ n)}\sigma_0\sigma_1}{\sigma_{1.23\ \cdots\ n}} \tag{42} $$

Let β_{01} be the path coefficient of the path from x_1 to x_0. Then by Wright's definition, in a two variable system,

$$ \beta_{01} = \frac{{}_1\sigma_0}{\sigma_0} = r_{01} \tag{43} $$

In a three variable system, $\beta_{01.2}$ may be defined as $\frac{1.2\sigma_0}{\sigma_0}$, *i.e.*, as the ratio of the part of the standard deviation of x_0 determined by x_1 *alone* to its total standard deviation.
Then

$$\left.\begin{aligned}\beta_{01.2} &= \frac{1.2\sigma_0}{\sigma_0} = \frac{r_{0(1.2)}\sigma_1}{\sigma_{1.2}}\\ \beta_{02.1} &= \frac{2.1\sigma_0}{\sigma_0} = \frac{r_{0(2.1)}\sigma_2}{\sigma_{2.1}}\end{aligned}\right\} \tag{44}$$

Let $d_{01.2}$ and $d_{02.1}$ be the corresponding coefficients of determination. Then

$$\left.\begin{aligned}d_{01.2} &= \frac{1.2{\sigma_0}^2}{{\sigma_0}^2} = \frac{{r^2}_{0(1.2)}{\sigma_1}^2}{{\sigma_{1.2}}^2}\\ d_{02.1} &= \frac{2.1{\sigma_0}^2}{{\sigma_0}^2} = \frac{{r^2}_{0(2.1)}{\sigma_2}^2}{{\sigma^2}_{2.1}}\end{aligned}\right\} \tag{45}$$

and there will be a portion of the variance of x_0 determined jointly by x_1 and x_2. Wright has shown that this

$$d_{0\overline{12}} = 2r_{12}\beta_{01.2}\beta_{02.1} \tag{46}$$

The sum of these coefficients of determination will be equal to the square of the multiple correlation coefficient.

In a system consisting of a criterion and n "independent" variables,

$$\left.\begin{aligned}\beta_{01.23\cdots n} &= \frac{1.23\cdots n\sigma_0}{\sigma_0} = \frac{r_{0(1.23\cdots n)}\sigma_1}{\sigma_{1.23\cdots n}}\\ \beta_{02.13\cdots n} &= \frac{2.134\cdots n\sigma_0}{\sigma_0} = \frac{r_{0(2.134\cdots n)}\sigma_2}{\sigma_{2.134\cdots n}}\\ &\vdots \qquad\qquad \vdots\\ \beta_{0n.123\cdots(n-1)} &= \frac{n.123\cdots(n-1)\sigma_0}{\sigma_0} = \frac{r_{0(n.123\cdots(n-1))}\sigma_n}{\sigma_{n.12\cdots(n-1)}}\end{aligned}\right\} \tag{47}$$

$$\left.\begin{aligned}d_{01.23\cdots n} &= {\beta^2}_{01.23\cdots n}\\ &\vdots \qquad\qquad \vdots\\ d_{0n.12\cdots(n-1)} &= {\beta^2}_{0n.12\cdots(n-1)}\end{aligned}\right\} \tag{48}$$

$$d_{0\overline{12}.34\cdots n} = 2r_{12}\beta_{01.23\cdots n}\beta_{02.13\cdots n} \tag{49}$$

$$d_{0\overline{ij}.23\cdots()\cdots()\cdots n} = 2r_{ij}\beta_{0i.12\cdots()\cdots n}\beta_{0j.12\cdots()\cdots()\cdots n} \tag{50}$$

where i takes any value from 1 to n, and j takes any value from 1 to n other then i, there being ${}_nC_2$ such terms.

There will also be terms of the nature of $d_{0\overline{123}.45\ldots n}$, $d_{0\overline{1234}.56\ldots n}$, $\ldots$, $d_{0\overline{1234\ldots(n-1)}.n}$, in all possible combinations. In problems involving more than three variables, the computation of these terms becomes exceedingly laborious. They are ordinarily of little interest in the study of the causal relationships in such problems, and are not necessary for the computation of the multiple correlation coefficient, which can be obtained from the path coefficients ($\beta_{01.23\ldots n}$, $\beta_{02.13\ldots n}$, etc.,) and the intercorrelations. Their calculation has been discussed by Wright.

From equations (44),

$$\left.\begin{aligned}\beta_{01.2} &= \frac{r_{0(1.2)}\sigma_1}{\sigma_{1.2}} = \frac{(r_{01}-r_{02}r_{12})\sigma_1}{\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{12}}} = \frac{r_{01}-r_{02}r_{12}}{1-r^2_{12}}\\ \beta_{02.1} &= \frac{r_{0(2.1)}\sigma_2}{\sigma_{2.1}} = \frac{(r_{02}-r_{01}r_{12})\sigma_2}{\sigma_2\sqrt{1-r^2_{12}}\sqrt{1-r^2_{12}}} = \frac{r_{02}-r_{01}r_{12}}{1-r^2_{12}}\end{aligned}\right\} \quad (51)$$

Similarly

$$\beta_{01.23} = \frac{r_{0(1.23)}\sigma_1}{\sigma_{1.23}} = \frac{r_{0(1.23)}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{12.3}}} = \frac{r_{0(1.3)}-r_{12.3}r_{0(2.3)}}{\sqrt{1-r^2_{13}}(1-r^2_{12.3})}$$

$$= \frac{\beta_{01.3}\sqrt{1-r^2_{13}}-\beta_{02.3}\sqrt{1-r^2_{23}}\,\beta_{21.3}\dfrac{\sqrt{1-r^2_{13}}}{\sqrt{1-r^2_{23}}}}{(1-r^2_{12.3})\sqrt{1-r^2_{13}}}$$

$$\beta_{01.23} = \frac{\beta_{01.3}-\beta_{02.3}\beta_{21.3}}{1-\beta_{21.3}\beta_{12.3}} \quad (52)$$

In general,

$$\beta_{01.23\ldots n} = \frac{r_{0(1.23\ldots n)}\sigma_1}{\sigma_{1.23\ldots n}} =$$

$$\frac{r_{0(1.23\ldots n)}}{\sqrt{1-r^2_{1n}}\sqrt{1-r^2_{1(n-1).2}}\ldots\sqrt{1-r^2_{12.3\ldots n}}}$$

$$= \frac{r_{0(1.3\ldots n)}-r_{12.3\ldots n}r_{0(2.3\ldots n)}}{\sqrt{1-r^2_{1n}}\sqrt{1-r^2_{1(n-1).2}}\ldots\sqrt{1-r^2_{12.3\ldots n}}}$$

$$\beta_{01.23\ldots n} = \frac{\beta_{01.34\ldots n}-\beta_{02.34\ldots n}\beta_{21.34\ldots n}}{1-\beta_{21.34\ldots n}\beta_{12.34\ldots n}} \quad (53)$$

It may be seen from formula (53) that the path coefficient $\beta_{01.23\ldots n}$ is identical with T. L. Kelley's partial regression coefficient in the

special (standard score) form of the multiple regression equation.* Hence, in a system involving an effect and several "independent" causes, the path coefficients of the paths from all the causes to the effect may be obtained simultaneously from the least squares solution of the standard-score multiple regression equation.†

E. L. Kelly[12] has pointed out that if one of the "independent" causes is itself the effect (in part) of other remote causes, we may set up this variable as a temporary criterion with the remote causes as the "independent" variables, and solve by means of the regression equation. Wright has shown that the path coefficient of the path from a remote cause to an effect is equal to the product of the path coefficient of the path from the remote cause to the immediate cause and that from the immediate cause to the effect, or to the sum of all such products, if the remote cause affects the effect through several immediate causes. He has also shown that if the causes do not combine by simple addition, but by multiplication or by more complex types of interaction, the path coefficients are *still* applicable as first approximations.

VII. Partial-multiple Correlation

In combining several series of "independent" observations for some specific purpose, it might be desirable to use as a criterion some partial function rather than a single variable. For example, we may desire to construct an educational test battery for the purpose of classifying pupils into groups on the basis of ability to do school work. The criterion group would necessarily be a sample of a wide range of school grades, the presumption being that those in the upper grades possessed the greater total ability. But school achievement is known to be a growing trait; hence our criterion should be that part of grade placement which is uncorrelated with chronological age, and each "independent" variable should be that part of the corresponding test score which is uncorrelated with chronological age. Since correlations with chronological age are known to be non-linear in large age

* Kelley was undoubtedly aware of the significance of this function when he stated[9] (p. 285), "It ($\beta_{01.2}$) is the weight or multiplying factor of z_1 when z_1 and z_2 are both used to estimate z_0."

† This fact was pointed out by E. L. Kelly[12] in connection with a discussion of Garrett's[13] RB functions. Neither Garrett nor E. L. Kelly has noted that these RB functions are identical with T. L. Kelley's β-regression coefficients.

ranges, it would be necessary to eliminate the squares and possibly the cubes of this variable as well as its first powers.

Let z_0 be the "dependent" variable (measured from the mean in terms of the standard deviation as a unit).

z_a, z_b, etc., to z_p, the eliminated variables.
z_1, z_2, etc., to z_n, the "independent" variables.
β_1, β_2, etc., to β_n, the corresponding regression weights.

(Then $\beta_1 = \beta_{(0.ab \cdots p)(1.ab \cdots p) \cdot (2.ab \cdots p)(3.ab \cdots p) \cdots (n.ab \cdots p)}$ in our former notation. The subscript 1 is sufficient to identify the β under consideration.)

Then $z_{0.ab \cdots p}$ will be the partial criterion. $z_{1.ab \cdots p}$, $z_{2.ab \cdots p}$, etc., to $z_{n.ab \cdots p}$ will be the partial "independent" variables. $\bar{z}_{0.ab \cdots p}$ will be the best estimate of $z_{0.ab \cdots p}$ which can be made from a knowledge of the partial "independent" variables. The problem is then to determine $\beta_1, \beta_2, \ldots \beta_n$ in such a manner that $\Sigma(z_{0.ab \cdots p} - \bar{z}_{0.ab \cdots p})^2$ shall be a minimum.*

$$\bar{z}_{0.ab \cdots p} = \beta_1 z_{1.ab \cdots p} + \beta_2 z_{2.ab \cdots p} + \cdots + \beta_n z_{n.ab \cdots p}$$

$$k^2_{(0.ab \cdots p) \cdot 12 \cdots n} = \frac{1}{N}\sum(z_{0.ab \cdots p} - \bar{z}_{0.ab \cdots p})^2$$

$$= \frac{1}{N}\sum(z_{0.ab \cdots p} - \beta_1 z_{1.ab \cdots p} - \beta_2 z_{2.ab \cdots p} - \cdots - \beta_n z_{n.ab \cdots p})^2$$

Differentiating with respect to β_1, and setting the first derivative equal to zero,

$$\frac{2}{N}\sum(z_{0.ab \cdots p} - \beta_1 z_{1.ab \cdots p} - \beta_2 z_{2.ab \cdots p} - \cdots - \beta_n z_{n.ab \cdots p})(-z_{1.ab \cdots p}) = 0$$

Differentiating with respect to each of the other β's and setting these derivatives equal to zero,

$$\frac{2}{N}\sum(z_{0.ab \cdots p} - \beta_1 z_{1.ab \cdots p} - \beta_2 z_{2.ab \cdots p} - \cdots - \beta_n z_{n.ab \cdots p})(-z_{2.ab \cdots p}) = 0$$

. . . . . .
. . . . . .
. . . . . .

* A solution of this problem by a method different from the one used here was given by Dr. Harold Hotelling in a paper entitled "Mixed Multiple-partial Correlation," which he read at a meeting of the San Francisco section of the American Mathematical Society at Berkeley Cal., Oct. 31, 1925. This paper was never published, but a brief report was printed in the *Bulletin* of the American Mathematical Society, Vol. XXXII, p. 98 (1926).

$$\frac{2}{N}\sum(z_{0.ab\ldots p} - \beta_1 z_{1.ab\ldots p} - \beta_2 z_{2.ab\ldots p} - \cdots - \beta_n z_{n.ab\ldots p})(-z_{n.ab\ldots p}) = 0$$

$$\vdots$$

Multiplying by the second factor, summing, and expressing in terms of correlation, we have the following normal equations:

$$\begin{aligned}
&r_{01.ab\ldots p} - \beta_1 - \beta_2 r_{12.ab\ldots p} - \cdots - \beta_n r_{1n.ab\ldots p} = 0\\
&r_{02.ab\ldots p} - \beta_1 r_{12.ab\ldots p} - \beta_2 - \cdots - \beta_r r_{2n.ab\ldots p} = 0\\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots\\
&r_{0n.ab\ldots p} - \beta_1 r_{1n.ab\ldots p} - \beta_2 r_{2n.ab\ldots p} - \cdots - \beta_n = 0
\end{aligned}$$

(Equations 54)

These equations may be solved simultaneously for the β's by the Doolittle method, the iteration method, or by determinants. It will be noted that these equations differ from those given by Kelley[9] (p. 295) for the ordinary multiple regression equation only in that the basic data are the partial correlation coefficients instead of the corresponding total correlations. The b-regression coefficients may be obtained from the relations,

$$b_1 = \beta_1\frac{\sigma_{0.ab\ldots p}}{\sigma_{1.ab\ldots p}},\ b_2 = \beta_2\frac{\sigma_{0.ab\ldots p}}{\sigma_{2.ab\ldots p}},\ \cdots b_n = \beta_n\frac{\sigma_{0.ab\ldots p}}{\sigma_{n.ab\ldots p}} \qquad (55)$$

The partial-multiple correlation coefficient is given by the formula

$$r_{0(12\ldots n)\cdot ab\ldots p} = [\beta_1 r_{01.ab\ldots p} + \beta_2 r_{02.ab\ldots p} + \cdots + \beta_n r_{0n.ab\ldots p}]^{\frac{1}{2}} \qquad (56)$$

Tappan[14] has discussed other methods for computing the partial-multiple correlation coefficient, which shorten the labor of computation when the regression coefficients are not required. Writing his equations in our notation,

$$r_{0(12\ldots n)\cdot ab\ldots p} = [1 - k^2_{01.23\ldots nab\ldots p}k^2_{02.34\ldots nab\ldots p}\cdots k^2_{0n.ab\ldots p}]^{\frac{1}{2}} \qquad (57)$$

$$r_{0(12\ldots n)\cdot ab\ldots p} = \left[1 - \frac{k^2_{0(12\ldots nab\ldots p)}}{k^2_{0(ab\ldots p)}}\right]^{\frac{1}{2}} \qquad (58)$$

VIII. Summary

1. The application of all forms of correlation to the solution of the problems of causation rests for its interpretation on the principle of the analysis of variance.

2. Partial correlation has been shown to be inadequate for solving most problems of causation.

3. Several systems of semi-partial correlation have been developed to meet certain of these deficiencies.

4. The relations of Wright's path coefficients and coefficients of determination to the principle of the analysis of variance, the second system of semi-partial correlation and the standard score form of the multiple regression equation have been pointed out.

5. A method for using a partial criterion and partial "independent" variables in the multiple regression equation has been discussed.

6. Other systems of semi-partial and semi-partial-multiple correlation could be devised to meet other needs. Those here presented were devised to aid in the solution of certain problems of psychology and education.

The writers are indebted to Dr. Harold Hotelling of Stanford University for reading considerable portions of this paper and offering several valuable suggestions.

References Cited

1. Mills, T. C.: "Statistical Methods." Henry Holt and Co., 1924, Chap. XII.
2. Tryon, R. C.: The Interpretation of the Correlation Coefficient. *Psychological Review*, 1929, pp. 419–445.
3. Pearson, K.: On a Form of Spurious Correlation Which May Arise When Indices Are Used, etc. *Proceedings Royal Society*, 1897, pp. 489–498.
4. Thomson, G. H., and Pintner, R.: Spurious Correlation and Relationships between Tests. *Journal Educational Psychology*, 1924, pp. 433–444.
5. Burks, B.: On the Inadequacy of the Partial and Multiple Correlation Technique. *Journal Educational Psychology*, 1926, pp. 532–540, 625–630.
6. Burks, B., and Kelley, T. L.: Statistical Hazards in Nature-nurture Investigations. *National Society for the Study of Education, Twenty-seventh Yearbook*, 1928, pp. 9–40.
7. Fisher, R. A.: "Statistical Methods for Research Workers." Oliver and Boyd, 1925, p. 172.
8. Yule, G. U.: "An Introduction to the Theory of Statistics." Charles Griffen and Co., 1919, Chap. XII.
9. Kelley, T. L.: "Statistical Method." The Macmillan Co., 1923.

10. Franzen, R.: A Comment on Partial Correlation." *Journal Educational Psychology*, 1928, pp. 194–197.
11. Wright, S.: Correlation and Causation. *Journal Agricultural Research*, 1921, pp. 557–585.
12. Kelley, E. L.: The Relationship between the Techniques of Partial Correlation and Path Coefficients. *Journal Educational Psychology*, 1929, pp. 119–125.
13. Garrett, H. E.: A Modification of Tolley and Ezekiel's method of Handling Multiple Correlation Problems. *Journal Educational Psychology*, 1928, pp. 45–49.
14. Tappan, M.: On Partial Multiple Correlation Coefficients in a Universe of Manifold Characteristics. *Biometrika*, 1927, pp. 39–44.

CIRCUMSTANTIALITY AS A FACTOR IN GUESSING ON TRUE-FALSE EXAMINATIONS

INA HILL BRINKMEIER AND NOEL KEYS[1]

University of California

The present experiment was inspired by the writers' convictions:

1. That the amount of guessing which takes place in the average true-false examination is much greater than is commonly realized, even by those who are being examined; and
2. That many of the statements in such examinations can be recognized as true from their form and nature, apart from any knowledge of the particular subject matter involved.

The fact that certain words or phrases in true-false statements often serve as "specific determiners," giving fairly dependable cues to the correct response has been noted by more than one writer.[2] For example, an overwhelming proportion of statements containing the words *all, always, only, no, never,* and *none* will be false, while statements qualified by *most, some, probably, may, often,* and the like, are true. Similarly, statements containing specific citations or enumerations, such as "——— was one of the ———," are usually true, while statements containing a cause or reason clause are more frequently false. Pupils soon become aware of these peculiarities of true-false tests, and there is little doubt that much of the practice-effect which gives so distinct an advantage to the "test-wise" pupil must be attributed to increasing sensitivity to cues of this sort.

The present article, however, is not concerned with such cue words and phrases, but with certain far subtler indications of a more general description. As previously stated, the writers have felt that such

---

[1] The compilation of the self-evident statements of List 1 and the practical conduct of this experiment were the work of Mrs. Brinkmeier; the devising of the factitious List 2 and the preparation of the present article, the work of Mr. Keys.

[2] Weidemann, C. C.: "How to Construct the True-false Examination." New York: Teachers College, Columbia University, Contributions to Education, No. 225, 1926.

Lee, Baldwin: Some Faults Common in Informal Tests Made by High School Teachers. *Educational Administration and Supervision,* Vol. XIV, February, 1928, pp. 105–113.

Brinkmeier, I. H.: "The Technique of Constructing Objective Examinations." University of California, Ph.D. Dissertation (unpublished), 1929.

indications are especially characteristic of true statements and that they are sufficient in themselves strongly to suggest the response "True," quite apart from any accurate knowledge of the points at issue.

In an attempt to secure more objective evidence of the existence and frequency of these indications, the writers obtained access, through the courtesy of Professors G. M. Ruch and G. A. Rice, to three hundred seventy-six objective examinations submitted in a nation-wide prize contest.[1] These examinations, assembled from thirty-one states, covered all of the principal departments of high school instruction, although slightly more than half dealt with English and the social studies. They included a total of 10,756 true-false items. Inspection of these last convinced the writers that more than one-fifth of the statements were of such a nature that their truth or falsity might be correctly inferred by an intelligent and "test-wise" reader, regardless of his knowledge of the subjects.[2] After eliminating the false statements in this list and all true items containing one or more cue words or phrases of the specific-determiner variety above mentioned, there remained one hundred statements the truth of which might be regarded as to some extent self-evident. From this number a random sampling of fifty was drawn for purposes of the present experiment.

An inspection of these fifty statements, as reproduced in List 1, following, will show that the characteristics which predispose one to the answer "True" vary widely from case to case. Some, such as Nos. 1 and 31, appear all but axiomatic. The student who would question the statement that "democracies need thinkers who will cooperate in the solution of problems," must be obtuse indeed. Similarly, in No. 38 it requires no knowledge of ionization to agree that "the theory of ionization offers a reasonable explanation of many facts but is not accepted as true by everyone"; the same might be said of almost any theory which one could name.

The truth of certain other statements, while a trifle less obvious, might yet be safely inferred from the most elementary reasoning. Examples are Nos. 4, 6, 24, and 47. Still others, such as Nos. 3, 21, and 28, would be strongly calculated to evoke the response of "True" on religious or patriotic grounds alone.

---

[1] Ruch, G. M., and Rice, G. A.: "Specimen Objective Examinations." Scott, Foresman and Co., 1929.

[2] Brinkmeier, I. H.: *Op. cit.*, Chaps. III and IV.

LIST 1.—FIFTY TRUE STATEMENTS FOUND IN ACTUAL EXAMINATIONS BUT REGARDED AS LIKELY BY REASON OF THEIR CIRCUMSTANTIAL NATURE TO BE MARKED "TRUE" WHETHER KNOWN TO THE PUPIL OR NOT. AFTER EACH STATEMENT IS SHOWN THE PER CENT OF "TRUE" RESPONSES RECEIVED FROM ONE HUNDRED EIGHTY HIGH SCHOOL AND UNIVERSITY STUDENTS

	PER CENT
1. Democracies need thinkers who will cooperate in the solution of problems	98.9
2. The movement of the earth around the sun, combined with the inclination of the earth's axis, gives us the seasons	97.2
3. Tennyson believed that a sinful man might still be a victor by overcoming his sin	96.1
4. Only when we deal with civil problems in the light of intelligence and understanding after investigation, can we hope for government in the interest of all the people	95.0
5. Slavery was such a vital question that it served as an incentive to literary activity	92.8
6. The adventures of the great Elizabethan seamen fired the imagination of other men of like adventure	90.6
7. The shortage of iodine in drinking water is a contributing cause of goitre	90.6
8. Early in life, Scott became interested in legends and in ruins of historical places	90.0
9. Abraham Lincoln's prose style shows a close acquaintance with the Bible	90.0
10. Southern writers by means of dialect and other touches of local color produced very realistic fiction	88.9
11. Radium, in spite of the fact that it is continually decomposing, is considered an element	88.9
12. Lamb's life was filled with tragedy and poverty, yet he enjoyed much that was cheerful and beautiful	88.3
13. "The Lady of the Lake" gives us many legends and quaint Highland customs	87.8
14. In the early period of our nation, laborers were too few and capital too scanty to exploit the wealth of our raw materials	85.8
15. The story of Beowulf lets us know something of the Anglo-Saxon ideals	85.5
16. The simple religious faith of Whittier found expression in many exquisite lyrics	84.4
17. Whittier has made the snows and the bees of his New England home famous	83.3
18. The "de Coverley Papers" reveal the English dress, manners, customs, tasks, and social, religious, and political beliefs of the early eighteenth century	83.3
19. The freed slave had civil and political rights in the United States more than a generation before the women	82.7
20. Representative government in England developed as a result of great charters of liberties which the kings were forced to grant	82.1
21. The Declaration of Independence states the belief that governments are instituted to secure the inalienable rights of man	81.1

22. With the close of the French and Indian War there came a distinct change in the behavior of England toward the colonies..... 81.1
23. Lowell, who has been regarded at home and abroad as a representative of the best American life and letters, stands for the power of literature to influence the thought and life of a nation..... 80.4
24. The rugged, inaccessible country of the Lake Katrine region prevented the Lowlanders from retaliating upon the plundering Highland clans. 80.0
25. Page's books dealing with stories of the Negroes and plantation days contain a note of romanticism..... 80.0
26. Aged persons and children are peculiarly susceptible to the respiratory diseases..... 78.9
27. The supporters of Hamilton's measures were a small minority of the total population, but common interests bound them in a compact, alert body, which enabled them to control the government..... 78.3
28. Throughout our Constitution runs the same high idealism concerning human equality as is announced in the Declaration of Independence 77.2
29. Sewall's diary portrays a vivid picture of the colonial life of his time... 76.5
30. In America, the group of western writers of the nineteenth century each had an unusual development of humor, democratic spirit, and a desire to present elemental facts of life..... 76.1
31. It is impossible to say in what proportion the motives of political protest, commercial adventure, economic pressure, and religious dissent were mixed in the emigrants of the seventeenth century..... 75.0
32. The first quarter of the nineteenth century saw the beginnings of a true literature in the department of poetry, fiction, and belles-lettres..... 74.4
33. The Colonial literary productions were impelled by the practical issues of the politics or theology of the time..... 74.0
34. In Virginia the growing of tobacco led to the occupation of large tracts of land and made impossible the establishment of the town with its local democratic features of meeting house and public school..... 73.3
35. The Jacksonian era saw the introduction of steam locomotion, reforms in criminal codes, in the curricula of schools, and in the organization of labor..... 73.3
36. Modern industrialism makes nations so interdependent that if the prosperity of one is injured from any cause, that of the others suffers in consequence..... 73.2
37. Johnson was a great scholar and the most important member of the famous Literary Club..... 72.6
38. The theory of ionization offers a reasonable explanation of many facts, but is not accepted as true by every one..... 71.7
39. Goldsmith's father was the model for the parson in the "Vicar of Wakefield"..... 71.5
40. The Renaissance aroused curiosity in the minds of the people of the Middle Ages..... 70.6
41. The early playhouses were built outside the walls of London..... 70.0
42. The value of the Erie Canal is not to be measured in tons actually carried 68.3
43. Samuel Richardson's early training as a novelist was gained through writing model love-letters for young ladies..... 67.2

44. The word "pageant" was first used to mean the platform upon which plays were performed 66.1
45. The history of the United States furnishes us with many examples of the peaceful settlement of international disputes 65.6
46. English aristocrats and authors of the Augustan period frequented coffee houses 65.6
47. The coming of the printing press led to the waning of folk-ballads 62.2
48. Coleridge was interested in a communistic colony to be established in America 56.7
49. The "Bachelor" of the "de Coverley Club" belonged to the Inner Temple, one of the most prominent lawyers' societies 56.4
50. One hundred years before the opening of Oklahoma for settlement, George Washington was inaugurated president 56.1

The majority of items in the list cannot, however, be so readily dismissed. The quality which they have most in common is perhaps best described as a certain circumstantiality of content and phrasing. The details are simply not the sort that a test-maker would be expected to invent. As one graduate student remarked when asked why he had marked statements sixteen and seventeen true although professing a total ignorance of Whittier, "If those statements were not true, they would not be in the test."

This holds true particularly of certain long statements in the list. As the writers expect to show elsewhere, there is a strong presumption that the longer statements in a true-false examination will be true ones. And when length takes the form of a cataloging of details, as in Nos. 18, 30, and 35, this presumption becomes all but a certainty. Such statements bear the stamp of the textbook all too plainly upon them. In them the testmaker appears to have abandoned the examining function for the didactic.

In order to obtain light upon the extent to which responses of "True" made to the above fifty items were based upon knowledge or understanding, the present writers prepared a second list of twenty-five statements of their own construction. These statements are reproduced in List 2, below. The items in this list are modeled closely upon statements chosen from List 1. The parallelism of statement A with statement 17, for example, will be at once apparent, as will the resemblance of D to 7, G to 50, J to 37, K and O to 20, L to 41, Q to 29, R to 18, S to 16, T to 4, W to 44, X to 39, and Y to 35, to mention but a few.

In each case, however, there is the striking difference that, whereas statements in List 1 are true, those of List 2 are false. The fifteen

items A to O inclusive range from the specious plausibilities of Riley's poem on "The Old Back Fence" (item A), and the deplorable shortcomings of a grass diet (item D), to such gross absurdities as the exploits of the Roman general, Cerberus, at Marathon (item I) or the choral arrangement of the Canterbury Tales (item E). All, however, share the common characteristic that they would assuredly be branded as false by one fully cognizant of the facts.

LIST 2.—TWENTY-FIVE FACTITIOUS STATEMENTS, FALSE OR REFERRING TO IMAGINARY PERSONS OR CIRCUMSTANCES BUT PHRASED WITH A CIRCUMSTANTIALITY SIMILAR TO THE FIFTY TRUE STATEMENTS OF LIST 1, TOGETHER WITH THE PER CENT OF HIGH SCHOOL AND UNIVERSITY STUDENTS REPLYING "TRUE" TO EACH

	Per Cent
A. Riley has made famous "The Old Back Fence" of his boyhood home.	86.1
B. The Antipodes are islands in the South Seas visited by Charles Darwin while collecting material for his "Origin of Species."	80.6
C. The combination of southern tradition with frontier life in the blue-grass region of Kentucky gave rise to a new and characteristically American form of literature	77.7
D. The lack of calcium in hay and grass causes the hoof-and-mouth disease.	66.7
E. "The Canterbury Tales" were among the first English poems to be set to music	62.8
F. Browning's best poetry was written in iambic hexameter	61.7
G. Caesar was crowned Emperor of the Holy Roman Empire two hundred years after the birth of Vergil	55.6
H. The addition of soluble aconite to a sodium or potassium salt produces a compound which acts as a powerful heart stimulant	53.7
I. Cerberus demonstrated his ability as a Roman general in the battle of Marathon	43.9
J. Samuel Endicott wrote the first literary work to receive country-wide recognition in America	42.2
K. The Renaissance was a result of the establishment of reformed churches in England and Germany	37.8
L. The Miracle plays were first produced before the Custom House in London	36.7
M. In the American colonies highway robbers were punished less severely than officials who embezzled public funds	31.1
N. The quarrel between Clay and Webster was fanned by the dislike of Clay's wife for New Englanders	27.8
O. Women secured the right to vote in the United States as a result of the World War	25.7
P. Before marriage the Scotch lassies of whom Burns wrote were accustomed, instead of an engagement ring, to wear around the neck a bright ribbon known as a "curchie."	88.3

Q. Thomas Burnham's accurate pictures of life were founded upon experiences and observations of human nature made in the course of his travels..... 84.8
R. "The Long Trail" paints a glowing picture of the courage and endurance of the men of "covered wagon" days..... 77.8
S. Marston's quiet country life is reflected in his literary efforts..... 74.4
T. The education of the masses to appreciate the philosophy of idealism should lead to the elimination of disputes between capital and labor.. 67.2
U. The sensitive nature of Shelley was moved to indignation by the revelation of the horrors of the African Slave trade..... 66.7
V. A cure for the stomach complaint known as gastritis has recently been discovered by the French physician, Bourdier..... 65.5
W. The brascombe was a type of short-sword worn by tragedians in the time of Shakespeare..... 61.7
X. Shakespeare's step-brother was the original and counterpart of Iago in "Othello."..... 59.4
Y. Under the rule of Sylvester villages grew apace, domestic animals multiplied, and there were even the beginnings of simple mining operations in what is now western Germany..... 57.6

The ten items P to Y differ slightly from the foregoing in that either the persons or the circumstances described are fictitious. One would search in vain for the literary productions of "Marston" or "Thomas Burnham"; history does not record the administration of the Hooverizing Sylvester; Shakespeare had no step-brother; the "brascombe" is found in no dictionary; and the charming custom of the Scotch lassies of statement P exists only in the imagination of the present experimenters. Only in the reference to Shelley is there the shadow of a likelihood, and this finds no substantiation in his biographies. Accordingly, whereas items A to O are clearly false, it is equally clear that items P to Y cannot be true. Of all alike, it may be said that they are marked by the same circumstantiality of form and content which characterizes statements of List 1.

In preparation for the present experiment the fifty items of List 1 and the twenty-five of List 2 were thrown together. To offset in part the preponderance of obviously true statements, the experimenters next selected from the examinations submitted in the Ruch-Rice contest twenty-five additional items regarded as obviously false. Typical of these last were such statements as: "Daniel Boone was one of our presidents"; "High winds help to extinguish forest fires"; and "Hatcheries are built to destroy fish." The entire one hundred items were then intermingled in chance order and mimeographed under the title of a "General Information Test."

This test was submitted to three classes of tenth, eleventh, and twelfth grade pupils at the University High School, Oakland, California, and to a group of one hundred senior and graduate students in the Department of Education at the University of California. The students were instructed both orally and by printed directions that some of the statements in the list were true and some false, but that the number of true and of false statements did not necessarily balance. Each statement was, therefore, to be marked true or false on its own merits. In case the student did not know the answer to a given statement he was instructed to guess, but any answer in which guessing was necessary was to be indicated by a question mark placed after the plus or minus sign. *The importance of this was emphasized. Students were also assured that their answers would in no way affect their class marks.* Believing the test to be a genuine information test, the students entered into it with zest, and to all appearances gave it their best attention and efforts. Twenty-five minutes was allowed for marking the statements, which proved ample for practically all members. After discarding the papers of those who failed to follow the directions and the few who did not finish, there remained a total of ninety-four papers from the high school and eighty-six from the university group.

The responses to each of the one hundred test items were then tallied separately, according to whether the statement was marked true or false, and whether guessing had been indicated by a question mark. Results from the high school and university groups were also separately computed.

In List 1 will be seen the fifty statements selected as more or less self-evidently true, arranged in descending order of the frequency of "True" replies received from the two groups combined. It will be noted that each of the fifty statements received a clear majority of "Trues," while in all but the last seven cases the proportion of "Trues" to "Falses" was more than two to one.

Considerations of space forbid including here the complete returns for each of the fifty items, but the summary may be seen in Table I, following. As might be expected, nearly four out of five of the total replies were "True" and approximately two times out of three this assertion was made with confidence. Of the items marked known, 81.2 per cent of the total replies were "True." While this proportion was naturally somewhat higher among the university group than among the high school students, it was very similar in both.

TABLE I.—RESPONSES OF NINETY-FOUR HIGH SCHOOL AND EIGHTY-SIX UNIVERSITY STUDENTS TO FIFTY TRUE-FALSE ITEMS ACTUALLY TRUE BUT CONSIDERED LIKELY BY REASON OF THEIR CIRCUMSTANTIAL NATURE TO BE MARKED "TRUE" WHETHER KNOWN TO THE STUDENT OR NOT

Group tested	Items marked known to be		Items marked guessed to be		Total marked[1]	
	True	False	True	False	True	False
High school	2696	713	916	370	3612	1083
University	2967	596	497	231	3464	827
Total	5663	1309	1413	601	7076	1910
Per cent	63.0	14.6	15.7	6.7	78.7	21.3

[1] In fourteen instances an item was left unmarked, which accounts for the failure of the totals to amount to 9000.

For our purposes, however, interest centers chiefly in the fact that among the answers frankly marked as guessed, the proportion of "Trues" was only slightly lower. Seventy per cent of the "guessed" replies were in the affirmative, and the high school group proved even more successful than the university. What was true of the totals held true to a remarkable extent of the individual statements as well. In the case of only one item among the fifty (No. 4) did more persons guess false than true.

We have, then, the striking fact that *students who believe that they know the answer to these statements are right only eighty-one per cent of the time, while those who admit they are guessing are correct in seventy per cent.* Such figures go far to substantiate the impression that statements of the type of List 1 bear too many surface indications of the reply expected, and are poorly adapted to distinguish between the well-informed and the merely shrewd pupil.

The objection may be raised, however, that the preponderance of true replies is not due to the nature of these statements, but to the fact that they are indeed true. There are many shades of knowledge, and it is altogether probable that many answers marked as guessed are in reality based upon vague or unconscious recollection of the facts.

It is at this point that the false statements of List 2 provide a most valuable check. It will be remembered that, quite apart from the fact that these statements are really false, the dice were to some extent loaded against their being marked true. Of the other seventy-five state-

TABLE II.—RESPONSES OF ONE HUNDRED EIGHTY HIGH SCHOOL AND UNIVERSITY STUDENTS TO TWENTY-FIVE FACTITIOUS TRUE-FALSE STATEMENTS ACTUALLY FALSE OR REFERRING TO IMAGINARY PERSONS OR CIRCUMSTANCES BUT PHRASED WITH A CIRCUMSTANTIALITY SIMILAR TO THE FIFTY TRUE STATEMENTS OF TABLE I

The Fifteen Statements A to O Inclusive Are Easily Recognizable as Contrary to Fact. The Ten Statements P to Y Are Concocted of Fictitious Persons or Imaginary Circumstances

Item	Times marked known to be		Times marked guessed to be		Total times marked[1]	
	True	False	True	False	True	False
A	98	12	57	13	155	25
B	88	22	57	13	145	35
C	109	33	30	7	139	40
D	54	24	66	36	120	60
E	79	53	34	14	113	67
F	49	35	62	34	111	69
G	42	58	58	22	100	80
H	23	22	72	62	95	84
I	27	68	52	33	79	101
J	36	44	40	60	76	104
K	46	107	22	5	68	112
L	30	80	36	34	66	114
M	35	94	21	30	56	124
N	17	81	33	49	50	130
O	40	131	6	2	46	133
Total A–O.....	773	864	646	414	1419	1278
P	75	10	84	11	159	21
Q	68	6	83	21	151	27
R	77	21	63	19	140	40
S	49	14	85	32	134	46
T	109	48	12	10	121	58
U	87	38	33	22	120	60
V	27	35	91	27	118	62
W	34	17	77	52	111	69
X	42	33	65	40	107	73
Y	36	39	65	37	101	76
Total P–Y.....	604	261	658	271	1262	532

[1] For several statements the number of true and false replies does not total one hundred eighty. This is due to the fact that in nine cases a statement was left unmarked.

ments in the test, fifty are true and but twenty-five false. Subjects as "test-wise" as the group here concerned are accustomed to find approximately one-half the statements in a true-false examination false, and to take account of this fact in making their replies. It is altogether probable, therefore, that, despite directions to the contrary, many would approach the statements from List 2 with the attitude of suspecting them to be false unless known to be otherwise. Notwithstanding this twofold handicap, we find from Tables II and III that *in the total replies to these false statements, the "Trues" outnumbered the "Falses" by approximately three to two.*

TABLE III.—RESPONSES OF NINETY-FOUR HIGH SCHOOL AND EIGHTY-SIX UNIVERSITY STUDENTS TO TWENTY-FIVE FACTITIOUS TRUE-FALSE STATEMENTS ACTUALLY FALSE OR REFERRING TO IMAGINARY PERSONS OR CIRCUMSTANCES BUT PHRASED WITH A CIRCUMSTANTIALITY SIMILAR TO THE FIFTY TRUE STATEMENTS OF LIST 1

The Fifteen Statements A to O Are Recognizably Contrary to Fact. The Ten Statements P to Y Are Concocted of Fictitious Persons or Imaginary Circumstances

	Items marked known to be		Items marked guessed to be		Total items marked	
	True	False	True	False	True	False
Items A–O (15):						
High school	451	376	396	184	847	560
University	322	488	250	230	572	718
Total A–O	773	864	646	414	1419	1278
Per cent	28.7	32.0	23.9	15.4	52.6	47.4
Items P–Y (10):						
High school	343	121	364	111	707	232
University	261	140	294	160	555	300
Total P–Y	604	261	658	271	1262	532
Per cent	33.7	14.6	36.6	15.1	70.3	29.7
Total A–Y (25)	1377	1125	1304	685	2681	1810
Per cent	30.7	25.1	29.0	15.2	59.7	40.3

More detailed inspection of Table II reveals the following facts:

1. Of the fifteen definitely false items A to O inclusive, eight are marked "True" more often than "False." In all, the "True" replies outnumber the "False" by eleven per cent.

2. Considering only the replies at which the subjects admitted they were guessing, twelve of the fifteen, or eighty per cent of the statements, are marked "True" more often than "False."

3. To the ten fictitious assertions P to Y, the "True" responses outnumber the "False" in every case by from thirty-three to six hundred fifty per cent.

4. Considering only the replies given as known without guessing, the "True" responses outnumber the "False" for eight out of the ten statements.

From Table III we derive the additional facts:

1. To the false items A to O, sixty per cent of the high school pupils and more than forty-four per cent of the university students replied "True."

2. Of the 1419 "True" responses made to these fifteen items, seven hundred seventy-three, or a clear majority, give no indication of guessing. With regard to particular items, forty per cent of all the high school and fifty per cent of the university group were prepared to mark as a fact which they knew without guessing that "the Antipodes are islands in the South Seas which were visited by Darwin while collecting material for his 'Origin of Species.'" Sixty-two per cent of all the high school and sixty per cent of all the university students were equally certain that Kentucky was the birthplace of a new and characteristically American form of literature, while fifty-two per cent of the high school and thirty-four per cent of the total university group "knew" that the "Canterbury Tales" were early set to music.

3. Of the replies to items A to O marked as guessed, the "True" responses outnumber the "False" by fifty-six per cent.

4. To the wholly imaginary assertions of items P to Y, the "True" responses are to the "False" in the ratio of seven to three, and nearly half of these replies are given without indication of guessing. Even the university group rated these statements "True" five hundred fifty-five times as compared with only three hundred times "False."

5. The proportion of "True" to "False" responses in the case of these last ten items averages about the same for answers given as known as for those marked as guessed. For example, eighty-eight per cent marked item P as describing a genuine Scotch custom, while reference to Table II shows that almost one-half of that number gave no indication of guessing.

6. Although it is extremely unlikely that any considerable number of persons taking this test had any accurate knowledge of either the truth or falsity of these ten fanciful assertions, yet forty-nine per cent of the high school and forty-seven per cent of the university students marked them as true or false without indication of guessing.

7. While superior knowledge enabled the university students to excel the high school pupils in the number of statements recognized as false, the general effect of circumstantiality of form in suggesting the response of "True" was evidently much the same for both groups. This appears most clearly in connection with items P to Y.

Conclusions

1. Guesswork is so universal and habitual in connection with the answering of true-false examinations that high school and university students are alike unable to distinguish which of their answers are founded upon actual information and which are merely the result of impressions gathered from the general nature and appearance of the statements. To twenty-five false but circumstantial statements, more than fifty-five per cent of the answers given as known without guessing were wrong. Any system of evaluating test results which expects pupils to distinguish between answers known and those merely guessed at is, therefore, likely to prove unsound.

2. Answers given as known are somewhat, but only moderately, more likely to be correct than are those indicated as the result of guessing. In the present study seventy-two per cent of the former were correct as compared with fifty-two per cent of the latter. (These figures apply to the seventy-five statements combined, and are obtained by pooling the totals of Tables I and III.)

3. Statements marked by the circumstantiality of content and phrasing which characterize Lists 1 and 2 are usually true and are apt to be recognized as such by students, even in the absence of any real knowledge of the facts. Of the 4003 cases in which students admitted they were guessing, seventy per cent of the answers to the true statements and sixty-six per cent of the answers to the false statements were marked "True."

4. Statements as obviously true as those of List 1 add little to the value of the test, since they do not discriminate successfully between the pupil who is well-informed and the one who is merely test-wise or shrewd at recognizing surface indications in the statements themselves. The chances are almost as great that these will be answered correctly by guesswork as by knowledge. In the present study, students who felt that they really knew the answers to the fifty true statements were right in only eighty-one per cent of the cases, whereas those who admitted they were guessing were correct in seventy per cent.

5. It follows that, in addition to being careful in the use of cue words or phrases which act as "specific determiners" of pupils' responses, test-makers will do well to scrutinize their true statements to make sure that these are not such as to be self-evident, merely by reason of their circumstantiality. Experience shows that it is difficult to invent comparable false statements after the fashion of List 2 to

offset circumstantial true statements of the type of List 1. Moreover, the positive suggestion effect conveyed by these false and fanciful assertions is evidently so strong that it would probably be pedagogically unwise to introduce many such into a test. Rather it would seem best to eliminate as far as possible the types of true statements which bear so many surface indications of their verity as to be of little value for purposes of measurement.

STUDIES IN HANDEDNESS: II. TESTING BIMANUAL HANDEDNESS

R. H. OJEMANN

State University of Iowa

In the first article of this series a method of testing unimanual handedness was described. In this article the results obtained by administering several bimanual tests to a large unselected group of subjects will be described. The group used in this investigation was the same as that used in the investigation reported in the first article. In this investigation also the testing was done in a separate room where the subject was not disturbed by school activities. One pupil at a time was summoned and tested.

The following four bimanual tests were used:

1. *Sweeping Test.*—The subject was given a full-sized broom and his attention was directed to several small blocks of wood (dimensions 4″ × 4″ × ¾″) placed in a semicircle about four feet away from a fifteen-inch circle drawn on the floor. He was then told to sweep the blocks into the circle, moving about as much as he wished. A record was made of the hand which rested lower on the broom during the sweeping and of the side of the body on which the broom was used.

2. *Raking Test.*—The subject was given a full-sized rake and asked to rake the several crumpled leaves of paper which were scattered about on the floor into a fifteen-inch circle drawn on the floor. He was told that he could move about as much as he wished. A record was made of the hand which rested lower on the rake and of the side of the body on which the rake was used.

3. *Shoveling Test.*—The subject was provided with a light shovel and told to "scoop" the paper leaves he had raked together into a box which stood nearby. A record was made of the hand which rested lower on the shovel and of the side of the body on which the shovel was used.

4. *Batting Test.*—*S* was provided with a ball-bat and told to stand about four yards from the experimenter. *E* directed *S* to bat the ball which *E* tossed to him. An indoor baseball was used. A record was made of the hand which rested nearer the batting end of the instrument and of the side of the body on which the bat was used.

Before the results obtained on the bimanual tests can be studied in detail, the method of scoring the tests must be explained. The customary procedure in scoring bimanual tests is as follows: If the subject holds the right hand nearer the business end of the instrument and does the work on the right side of the body, a score of *R* is given. In sweeping, for example, if the subject places the right hand lower

on the broom and does the work on the right side of the body, he is classified as a right-handed sweeper. On the other hand, if the subject places the left hand lower on the broom and does the work on the left side of the body, he is designated as a left-handed sweeper.

This scheme of classification was used by Rife[1] and by Downey[2] in their studies of bimanual handedness. Both of these investigators found, however, that the majority of the human race sweep left-handed. In this investigation similar results were obtained. In the unselected group three hundred sixty-three of the five hundred eighteen subjects were found to be left-handed sweepers according to the customary classification. Similarly, three hundred seventy-two of the unselected group were left-handed in raking, and three hundred ninety-two were left-handed in shoveling. On the other hand, only twenty-four were left-handed in batting.

These results indicate that something must be wrong with the customary classification as applied to the sweeping, raking, and shoveling tests. A classification according to which seventy per cent of an unselected population-group are left-handed in sweeping, in raking, or in shoveling cannot be very serviceable. For this reason the customary classification as applied to sweeping, raking, and shoveling will be reversed, and the handedness in the bimanual activities will be studied in terms of the new classification.

Before turning to a study of the results obtained when the new classification is adopted, it is necessary to describe the way in which the scores were treated in those cases in which the subject held one hand nearer the business end of the instrument but did the work on the opposite side of the body, and in those cases in which the subject performed the activity in the *R* or *L* positions with equal ease. Cases of the former type were designated by the score *Cr* (crossed). Cases of the latter type in which the subject performed the activity in the *R* or *L* positions with equal ease, or in which the subject performed the activity in the two possible *Cr* positions with equal ease, were given a score of *A*.

The results on the bimanual handedness tests will be analyzed from two points of view. In the first place, the relation of the scores

---

[1] Rife, J. M.: Types of Dextrality. *Psychological Review*, Vol. XXIX, 1922, pp. 474–480.

[2] Downey, J. E.: Types of Dextrality and Their Implications. *American Journal of Psychology*, Vol. XXXVIII, 1927, pp. 317–367.

on the four bimanual tests to one another will be studied. In the second place, the relation of the results of the bimanual tests to unimanual handedness will be considered.

Of the five hundred eighteen cases composing the unselected group, all but seventy-three made consistent scores on the sweeping, raking, and shoveling tests. That is, ninety individuals made a score of *L* (new classification) on these three tests, and three hundred fifty-five made a score of *R* on the three tests. The seventy-three cases in which consistent scores were not made on the three tests are tabulated in Table I. The data in Table I show, for example, that there were twenty-five subjects who made a score of *L* in sweeping, *R* in raking and in shoveling.

The meaning of the close agreement between the scores on these three tests is obscured somewhat by the fact that, with few exceptions, there are practically only two possible scores, *R* or *L*, on each test. As a result of pure chance, therefore, complete agreement between the scores on all three tests would be obtained in twenty-five per cent of the cases. Since complete agreement was obtained in 85.91 per cent of the cases it is evident that the three tests measure some form of bimanual handedness.

TABLE I.—SCORES IN BIMANUAL TESTS FOR THE SEVENTY-THREE SUBJECTS IN THE UNSELECTED GROUP WHO FAILED TO MAKE CONSISTENT SCORES ON THE SWEEPING, RAKING, AND SHOVELING TESTS

SCORES ON THE SWEEPING, RAKING, AND SHOVELING TESTS	NUMBER
L in sweeping, *R* in raking and in shoveling	25
R in sweeping, *L* in raking and in shoveling	5
Cr in sweeping, *L* in raking and in shoveling	1
Cr in sweeping, *R* in raking and in shoveling	2
L in raking, *R* in sweeping and in shoveling	17
R in raking *L* in sweeping and in shoveling	6
A in raking, *R* in sweeping and in shoveling	2
R in shoveling, *L* in sweeping and in raking	11
L in shoveling, *R* in sweeping and in raking	4
Total	73

L Sweeping, raking & shoveling 90
R " " " 355

The relationship between the scores on the sweeping, on the raking, and on the shoveling tests and the scores on the batting test is shown by the data presented in Table II. The number of left-handed batters is much smaller than the number of individuals who are left-handed in sweeping, raking, or shoveling. The data

also show that ten of the twenty-four left-handed batters are left-handed in sweeping, eleven are left-handed in raking, and eleven are left-handed in shoveling. The relationship between the scores on the batting test and the scores on either the sweeping, raking, or shoveling tests is not very close. The number of left-handed batters in an unselected group is much smaller than the number of left-handed sweepers. Also, only about one-half of the left-handed batters are left-handed in sweeping. This seems to indicate that there is a difference between the batting activity and the type of activity involved in sweeping, in raking, or in shoveling. A discussion of the probable nature of this difference will be presented as soon as the relation of bimanual handedness to unimanual handedness has been considered.

TABLE II.—RELATIONSHIP BETWEEN THE SCORES IN BATTING AND THE OTHER BIMANUAL TESTS

Scores in sweeping, raking, or shoveling	Number	Score in batting		
		R	*L*	*Cr*
Scoring *R* in sweeping	363	309	14	40
Scoring *L* in sweeping	152	135	10	7
Scoring *Cr* in sweeping	3	2	0	1
Scoring *R* in raking	372	316	13	43
Scoring *L* in raking	144	126	11	5
Scoring *Cr* in raking	0	0	0	0
Scoring *A* in raking	2	2	0	0
Scoring *R* in shoveling	392	336	13	43
Scoring *L* in shoveling	126	110	11	5
Scoring *Cr* in shoveling	0	0	0	0

In studying the relationship of bimanual handedness to unimanual handedness, the scores on the sweeping, raking, and shoveling tests will be combined into a single score since it has been shown that there is a close agreement between the results on these three tests. The combined score for a given subject will be taken as the score made on two of the three tests. Thus if a subject made a score of *R* on

the sweeping test and a score of *L* on the raking and shoveling tests, the combined score would be *L*.

The data showing the relationship of the scores on the bimanual tests to unimanual handedness are presented in Table III. The figures in the first column refer to the combined scores on the unimanual handedness tests. The figures in the second column show the number of subjects in the unselected group whose scores in the unimanual handedness tests were of a given magnitude. The figures in the third column indicate the number of subjects who made a

TABLE III.—RELATION BETWEEN THE SCORES ON BIMANUAL TESTS AND UNIMANUAL HANDEDNESS

Unimanual handedness		Handedness in sweeping, raking, and shoveling			Handedness in batting			
Score	Freq. in uns. group	*R*	*L*	Proportion of *L*	*R*	*L*	*Cr*	Proportion of *L*
−16	1		1	1.00		1		1.00
−13	6		6	1.00	1	3	2	.50
−10	10	1	9	.90	6	4		.40
− 7	8	1	7	.87	4	3	1	.37
− 4	6	1	5	.83	5		1	.00
− 1	3	1	2	.67	3			.00
2	16	10	6	.37	14	1	1	.06
5	39	32	7	.18	32	4	3	.10
8	146	116	30	.21	126	1	19	.07
11	157	129	28	.18	140	4	13	.03
14	86	62	24	.28	80	2	4	.02
17	31	23	8	.26	27	0	4	.00
20	8	8	0	.00	7	1		.13
23	1	1	0	.00	1	0		.00

combined score of *R* on the sweeping, raking, and shoveling tests. The figures in the fourth column indicate the number of subjects making a combined score of *L* in the three tests. The figures in the fifth column show what proportion of the subjects in a given interval scored left on the sweeping, raking, and shoveling tests. For example, there were six individuals in the unselected group who made a score on the unimanual handedness tests falling in the interval negative

thirteen up to but not including negative ten. Six of these subjects made a combined score of *L* on the sweeping, raking, and shoveling tests. The proportion making a score of *L* is, therefore, 1.00. The figures in the last four columns of Table III can now be easily interpreted by the reader.

The data in Table III are presented graphically in Fig. 1. The units on the horizontal axis represent unimanual handedness scores. The units on the vertical axis represent the proportion making a score of *L*. A study of Fig. 1 shows that handedness in sweeping, raking, and shoveling and handedness in batting vary directly with handedness in unimanual activities. A very high proportion of those scoring *L* in unimanual activities also scored *L* in sweeping, raking, and shoveling. In the region of ambidexterity in unimanual handedness the proportion drops. As the right-hand region is entered the proportion drops still more, and remains low throughout the region of right-handedness in unimanual activities. The curve showing the proportion making a score of *L* in batting shows similar changes, but drops much more rapidly. Beginning with the negative four group the proportion remains very low.

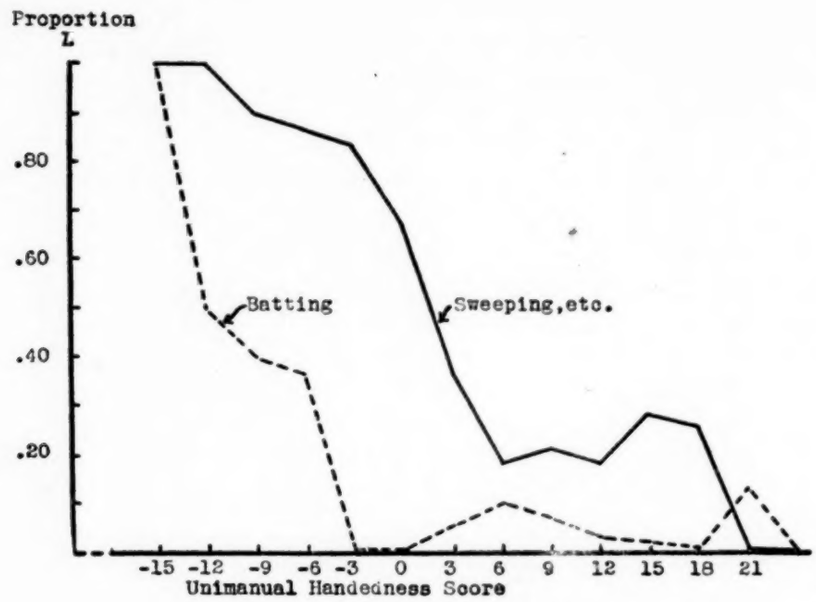


Fig.1 Relation Between Unimanual and Bimanual Handedness

The data show that handedness in bimanual activities is related to unimanual handedness but that the relationship varies somewhat

with the type of activity tested. The scores on the batting test show a behavior different from that shown by the scores on the sweeping, raking, and shoveling tests. The curve for the scores on the sweeping, raking, and shoveling tests drops gradually as one passes from the left-hand end of the unimanual handedness scale to the right-hand region. The curve for the scores on the batting test drops very rapidly. From the results it appears that batting is an activity of such a type that left-handed individuals can easily learn to perform it in the right-handed way, or that there are stronger environmental influences operating to favor right-handed batting than is the case with such activities as sweeping, raking, and shoveling. One possible explanation for the fact that the proportion of left-handedness in sweeping, etc., is so much higher for the group as a whole than the proportion of left-handedness in unimanual activities, is that these activities involve the two hands and that it is, therefore, not as difficult for a left-handed individual to learn to perform a bimanual activity in the right-handed way as it is to learn to perform a unimanual activity in the right-handed way. In a bimanual activity the two hands must each develop some degree of skill. Whether a given individual will learn to perform the activity in the right-handed or left-handed way will depend more upon factors other than the handedness of the individual than is true for unimanual activities. It also appears that the way an individual uses a bat is more conspicuous than the way the broom or shovel is used. Handedness in batting might, therefore, be more subject to social influences, and this condition, plus the probable fact that it is less difficult for a left-handed individual to learn to use a bat right-handed than it is to learn to perform a unimanual activity with the right hand, offers a possible explanation for the results obtained on the batting test. However, the foregoing explanations are only hypotheses and must be tested experimentally before any considerable weight is given them.

Summary

1. The customary classification of handedness as applied to such activities as sweeping, raking, and shoveling is in error. According to the customary classification, the majority of the human race is left-handed in these activities, and bimanual handedness varies inversely with unimanual handedness. When the customary classification is reversed, the majority of individuals is designated as right-

handed and the direct relationship of bimanual handedness to unimanual handedness is at once apparent.

2. The handedness in bimanual activities varies somewhat with the nature of the activity tested. The handedness in such an activity as sweeping is more closely related to the handedness in such activities as raking and shoveling than to the handedness in such an activity as batting.

STORY COMPLETION TESTS

GRACE H. KENT AND E. FRANCES WELLS

From Danvers State Hospital, Hathorne, Mass.

These two tests are based in principle upon the well-known Trabue Completion Tests.[1]

Whether a test be intended for school children or for clinical subjects, it is almost axiomatic that the material used should be of sufficient interest to hold the attention of the subject. Other things being equal, a one-page story is more interesting than is a page of detached sentences. But a child should not be required to leave a story unfinished, and for this reason a story would be unsuitable material for a test having a time limit. These tests were intended for standardization without time limit, and therefore it was possible to take advantage of the interest which a story naturally stimulates.

The selection of material for any test which is to be used both for children and for adults presents a special problem. It is essential that the subject-matter be within the grasp of children at the lower age levels, and yet it should be free from any puerile content such as might tend to make an adult or adolescent subject feel foolish. In view of the inherent difficulty of meeting these conflicting requirements, it seemed safer to fall back upon classics than to employ original literary material.

After considerable study of fables and anecdotes, the fables "Lion and Mouse" and "The Fir Tree" were adopted, largely on the ground that each can be reduced to one-page length without omitting anything essential. Following are the two stories in the form used, with blank spaces which are indicated by numbers in parentheses.

Name Only one word in each space.

A great lion (1) sleeping in the woods. A little mouse ran (2) him and wakened him. The (3) sprang (4) and caught the (5) under (6) paw. The mouse said: "If you will (7) me go, some day (8) will (9) useful to (10)." The (11) laughed and answered: "You (12) too (13) to be (14) any use (15) me; but (16) will let you (17) , because lions (18) generous."

A few days (19) this, the lion was (20) in a snare. He roared (21) rage, and all the (22) in the woods could (23) him. The little (24) came running to (25) place, to find (26) the reason

[1] Trabue, M. R.: "Completion Language Test Scales." Columbia Contributions to Education, 1916.

(27) all this noise. The mouse (28) to the lion: "I promised you (29) some (30) I would be (31) to you. My (32) are very sharp, and I (33) cut the cords of this (34) . I will (35) you free."

So the (36) gnawed at the heavy (37) which held the lion's (38) . Finally the cords were (39) through, and the (40) was set (41) .

Name Only one word in each space.

A small fir (1) grew in (2) woods. The tree said: "No one cares (3) me, because (4) needles are (5) so beautiful (6) the leaves of the other (7) . If only I had (8) of gold, people would like (9) best of (10) ."

The (11) morning the wish (12) true, and the tree (13) leaves of pure (14) . But (15) soon (16) man picked off (17) the leaves, and carried (18) away in a bag.

Then the tree (19) for leaves of glass, and (20) wish also came (21) . The (22) leaves sparkled in the (23) , and the (24) was very (25) . But after a while (26) storm arose, and the (27) shook the trees. All the pretty glass (28) were (29) off.

After that (30) tree wished for green (31) , like those of the (32) trees. But (33) this wish was granted, a goat (34) along and ate off (35) leaf.

Then the tree (36) : "My (37) were the best, because I could (38) them all the (39) . How I wish (40) had my (41) needles back!" The next (42) this (43) came true, and the fir tree laughed (44) joy.

Method of Standardization

Preliminary standardization of these tests was accomplished as a by-product, by a method already described.[1]

A battery of seven short tests, prepared for group standardization, was given to about one thousand children in the public schools of Lynn, Massachusetts. The children were from Grades III to IX, and their ages ranged from eight to fourteen. The battery included a new sentence completion test, "Woodworth-Wells Hard Directions,"[2] and a revised edition of five Kent-Shakow group tests. *Sentence Completion*, the only one having a time limit, was presented first;

---

[1] Kent, G. H., and Shakow, D.: Group Tests for Clinical Studies. *Journal of Genetic Psychology*, Vol. XXXV, 1928, pp. 595–618. (See p. 611.)
Kent, G. H.: "Battery of Seven Tests."

[2] Woodworth, R. S., and Wells, F. L.: Association Tests. *Psychological Monographs*, No. 5, 1911.

after which the other six tests were given out together, in such a way as to permit each child to set his own pace. The forms[1] of the *Story Completion* tests were held in reserve, and later were presented individually to those children who completed the battery early in the period and who therefore had time for extra work. The story tests thus furnished a convenient means of keeping the rapid workers busy while the slow workers were still occupied with the battery.

In Grade III preference was given to "Lion and Mouse," as being the easier test. In other grades the number given out was kept approximately equal for the two tests. "Lion and Mouse" was completed by four hundred forty-four children, and "Fir Tree" by four hundred twenty-eight children. There were six hundred twenty-two children who completed the battery in time to perform at least one of the extra tests, and two hundred fifty of these—obviously the very rapid workers—had time to complete both. Thus the children who took part in these tests do not constitute a representative group of school children, because the slow workers were entirely excluded. For this reason we are not using the records for independent standardization of the tests, but merely for deriving such norms as can be obtained by correlation with the battery.

Evaluation of Results

The chief purpose of these tests is to offer the subject an opportunity to produce a logically coherent story. They were not intended as tests of memory or verbal association, nor primarily as tests of ability to use language correctly.

A scoring system should be sufficiently flexible to meet the demands of the material without being so complex as to be unwieldy. In a test having forty items which are not arranged in order of difficulty, the weighting of values for individual items would increase the difficulty of scoring far more than it would increase the accuracy of results. It seems best, therefore, to allow the same credit for all items, although they vary widely in difficulty.

---

[1] Mimeographed copies were used for the blank forms, made by stencils prepared on a typewriter having pica type. The lines were triple-spaced. In order to allow ample room for writing the longer words, without wasting the page, the blank spaces were graduated roughly according to the length of the words to be supplied. (Sample copies of the forms will be furnished on application.—G. H. K.)

A preliminary study of the records revealed a considerable variety of responses for almost any given item. In many instances two or more of the words seemed equally to merit full credit; other responses seemed to merit partial credit, although they were less satisfactory; and still others were unquestionably inadequate. Finding it impossible to draw a sharp line between "right" and "wrong" responses without being arbitrary, we decided to allow two points for each response that is wholly satisfactory and one point for each inferior response that makes fairly good sense.

It is frequently difficult to decide whether a given response belongs in the superior or inferior group. In order to reduce the effect of personal equation, a few general rules have been formulated to serve as a guide in scoring. Although these rules still leave much to individual judgment and therefore do not wholly eliminate the personal factor, it seems better to admit this source of error than to introduce the errors which an inflexible scoring system involves.

General Rules for Scoring

Four criteria are recognized, as follows:

1. Appropriateness of the response. To receive any credit, it is essential that the response makes good sense in relation to the context.
2. Clearness with which the appropriate meaning is conveyed.
3. Grammatical correctness of the construction.
4. Observance of the instructions—to write only one word in a space.

Two-point credit is allowed for any response which satisfies all four of these requirements.

One-point credit is given for a response which meets the first condition but which violates any one of the other three. Thus:

(*a*) A word which is grammatically incorrect and awkward, but which conveys the appropriate meaning clearly.

(*b*) A word which does not convey the meaning clearly, but which is grammatically correct.

(*c*) A two or three word response which meets the first three conditions.

No credit is allowed for words which do not make good sense in the context, nor for grammatically incorrect words which do not convey the meaning clearly.

No credit is deducted for incorrect spelling, provided that the meaning be clear in spite of the error.

The records were scored according to these principles. The points for full-credit responses were totaled and recorded as the *First Score.* The points for half-credit responses were then added to the first score, and the total number of credits was recorded as the *Second Score.*

The words for which we have given credit are listed in the keys, the items being referred to by number. Responses which receive full credit are in bold-face type, followed by the half-credit responses in plain type. These lists are purely empirical, and are not exhaustive. They were developed in the process of scoring the records, each new response that was found being added to the list. The individual examiner will doubtless have occasion to add other words from time to time according to the demands of the results.

Key for Lion and Mouse

(1) **Lay, was**—is, lie. (2) **Across, over**—about, around, before, beside, by, into, near, on, onto, to, towards, up, upon, up to. (3) **Lion**—cat. (4) **Forward, quickly, suddenly, up**—after, at, at him, for, forth, quick, upon. (5) **Mouse**—rascal, rat. (6) **His**—her, its, the. (7) **Let**—leave. (8) **I.** (9) **Be.** (10) **You.** (11) **Lion.** (12) **Are.** (13) **Little, small.** (14) **Of.** (15) **To**—too. (16) **I.** (17) **Go, live**—free, loose. (18) **Are**—were. (19) **After.** (20) **Captured, caught, trapped**—tangled, tied. (21) **In, with**—for, in a. (22) **Animals, creatures, hunters, people**—birds, lions, mice, neighbors, others, rats, squirrels, trees. (23) **Hear**—not help. (24) **Mouse**—rat. (25) **The**—his, that, this. (26) **Out**—what. (27) **For**—of, was for. (28) **Said**—yelled. (29) **That**—once. (30) **Day, time**—some day. (31) **Useful, helpful**—a help, generous, good, kind, of service, of use. (32) **Teeth**—little teeth. (33) **Can, will**—could, shall, 'll. (34) **Net, snare, trap**—cage, rope. (35) **Set**—cut, get, let, make. (36) **Mouse**—rat. (37) **Cord, cords, net, rope, ropes**—snare, trap. (38) **Body, feet, legs, paws**—foot, head, paw. (39) **Cut, gnawed**—all cut, bitten, broken, eaten, torn. (40) **Lion.** (41) **Free**—loose.

Key for Fir Tree

(1) **Tree**—once. (2) **The**—a, some. (3) **About, for.** (4) **My**—prickly, the. (5) **Not**—never, n't. (6) **As**—like. (7) **Trees**—tree. (8) **Leaves**—needles, some, those. (9) **Me**—mine, them. (10) **All**—course. (11) **Following, next**—same. (12) **Came**—come. (13) **Had**—bore, grew, has. (14) **Gold.** (15) **Pretty, very**—quite, then. (16) **A, some**—the. (17) **All**—some of. (18) **Them**—it. (19) **Longed, wished, yearned**—asked, begged, sighed, cried, wish. (20) **That, this**—her, his, its, the. (21) **True.** (22) **Glass**—beautiful, new, pretty, tree's. (23) **Sun, sunlight, sunshine**—forest, light, moonlight, woods. (24) **Fir, tree**—glass, leaves, others, sight. (25) **Beautiful, happy, pretty, proud**—bright, glad, handsome, jealous. (26) **A**—the. (27) **Wind**—storm. (28) **Leaves**—leaf, leave, needles. (29) **Blown, broken, shaken**—broke, fallen, falling,

knocked, shook, taken, torn. (30) **The.** (31) **Leaves**—leaf. (32) **Ordinary, other**—big, elm, maple, pretty. (33) **After, when**—although, as soon as. (34) **Came**—went. (35) **Every**—each. (36) **Said, sighed, thought**—cried, mourned. (37) **Needles**—leaves. (38) **Have, keep**—own, wear. (39) **Time, year.** (40) **I.** (41) **Fir, old, own**—beautiful, dear, green, lovely, nice, pointed, pretty, sharp. (42) **Day, morning**—minute, summer, time, year. (43) **Wish**—also, one. (44) **For, with**—in, of.

Development of Norms

The battery yielded seven independent mental age ratings for each child, and the median of seven was accepted as the final rating. The *Story Completion* records were grouped according to these ratings instead of according to the ages of the children, and the scores for each age-level were arranged in serial order. At any given level the actual ages of the children varied widely. At the ten-year level, for example, the chronological ages ranged all the way from eight to fourteen.

Tables I and II show the median scores for each level, with the range of variation. The medians are shown also in graphic form, in Figs. 1 and 2. The solid lines show the first score medians, and the broken lines the medians for the second score.

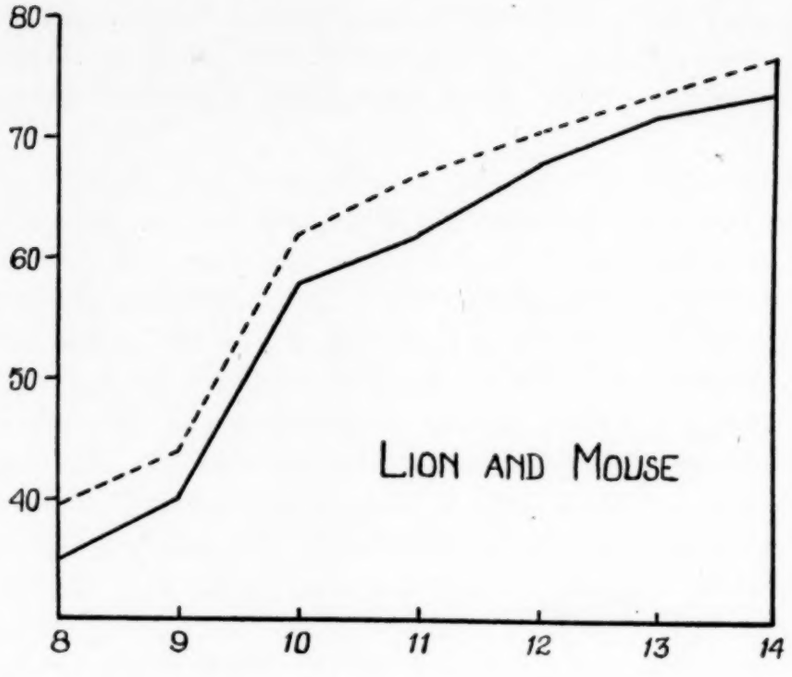
It will be noted that the median scores at the nine-year level are only slightly above those of the eight-year level, which indicates that the tests are rather too difficult at nine years. It is from ten to fourteen years that they are clearly discriminative, and we do not expect them to be of any considerable use outside of this range.

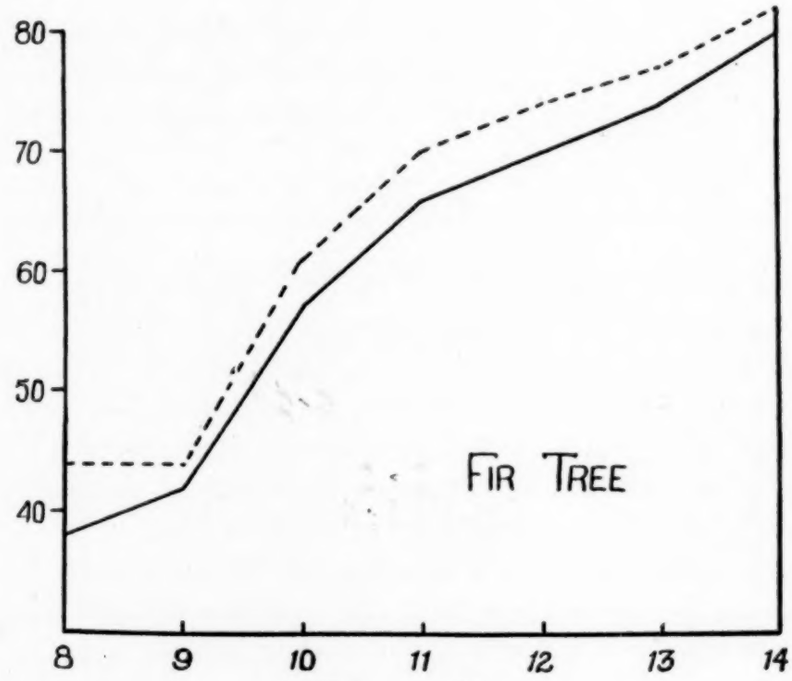
The norms which we are using, derived from the medians, are shown in Table III. The curves have not been altered, but they have been extended at the upper end so as to give a rating of fourteen-plus for a score that is clearly too high to be counted as fourteen. In each instance the actual median for a given level is included within the range for the corresponding age.

We use the second score in preference to the first score; however, the norms by the first score are given in full for the convenience of any who may prefer this method of scoring.

Serviceability of the Tests

These tests have seen some service in Danvers State Hospital, both in the Child Guidance Clinic and in the study of adult hospital


80
70
60
50
40
LION AND MOUSE
8
9
10
11
12
13
14


80
70
60
50
40
FIR TREE
8
9
10
11
12
13
14

FIGS. 1 and 2.

patients. They appear to be exceptionally attractive tests and to furnish a task of absorbing interest. Inadequate as the norms are, we have found the tests useful in their present state, and it is on this ground that we offer them for the use of other clinics. Written tests which have no time limit are at a premium, especially for individual studies.

Standardization of tests on a large scale is a project for school systems rather than for state hospitals, and we are not prepared to carry the development of these tests any further. If they should be used as group tests for classification of school children, material might be available for adequate standardization. In the meantime we can safely recommend them for use in school clinics as individual tests. They are offered, of course, merely as furnishing one more element in an examination, preferably to be used as one unit of a battery. It is obvious that no subject should be given a rating based wholly upon so simple a test, even properly standardized. (In this hospital we do not make a practice of rating a subject by any single test, whether simple or composite.)

The two tests are intended to be used singly rather than together. Either is probably sufficient to offer a rough measure of whatever aptitude is measured by completion tests. "Lion and Mouse" is to be preferred for children under twelve years of age or for adults who rate below twelve, and "Fir Tree" is recommended for any subject who may reasonably be expected to rate upward of thirteen.

Summary

Two simple and well-known fables have been arranged for use as story completion tests. In each test approximately forty words were deleted, each being represented by a blank space. The subject is expected to fill in the words required to complete the story. Two-point credit is allowed for every satisfactory word supplied, and one-point credit for an inferior response which makes good sense in the context. The final score is the sum of these credits.

The tests have been tentatively standardized upon over four hundred children, by correlation with the median ratings which these children achieved in a battery of seven short units. The age-curves required no essential correction.

The tests are given without time limit. They are intended primarily for use in individual examination of clinical subjects, but can

be used also as group tests. They are discriminative from ten to fourteen years.

TABLE I.—LION AND MOUSE

Age level	First score		Second score		Number of cases
	Median	Range	Median	Range	
8	35	6–62	39.5	6–66	22
9	40	8–64	44	9–67	39
10	58	28–78	62	33–79	57
11	62	42–74	67	44–78	77
12	68	34–80	70.5	39–80	127
13	72	54–82	74	58–82	107
14	74	66–78	77	69–80	15

TABLE II.—FIR TREE

Age level	First score		Second score		Number of cases
	Median	Range	Median	Range	
8	38	34–52	44	35–52	5
9	42	2–74	44	2–75	26
10	57	22–84	61	25–84	50
11	66	34–74	70	37–77	85
12	70	24–88	74	28–88	134
13	74	46–86	77	51–87	112
14	80	56–86	82	61–87	16

TABLE III.—TENTATIVE NORMS

Age level	Lion and Mouse		Fir Tree	
	First score	Second score	First score	Second score
10	52–58	56–64	52–60	56–65
11	60–64	65–68	62–66	66–71
12	66–68	69–72	68–72	72–75
13	70–72	73–76	74–76	76–79
14	74–78	77–79	78–82	80–84
14+	80–82	80–82	84–88	85–88

SEX DIFFERENCES IN ALGEBRAIC ABILITY

GLENN R. PEASE

College of the Pacific

The author has no desire to review the literature respecting sex differences in algebraic ability; in fact he knows of but little which could be reviewed. His only interest is in adding some data that seem worthy of consideration in connection with this subject.

After comparing the scores of three hundred sixty-four boys and three hundred twenty-one girls who were given the forty element I.E.R. test to quadratics, Thorndike states: "The evidence supports the conclusion that the sexes are of approximately equal ability." ("Psychology of Algebra," p. 414.) After noting Thorndike's conclusion, it seemed worth while to the author to study sex differences in algebraic ability in connection with a rather extended research carried on for an entirely different purpose. The present report is really an appendix to a study printed in *The Mathematics Teacher* of May, 1929, under the title, An Analysis of the Learning-units in *N* Processes in Algebra. If the reader desires a more detailed statement of the manner of setting up the tests, and the carrying out of the test program than is included here, it can be found in that article.

The field of abstract whole-number problems in addition, subtraction, multiplication, division, parentheses, and algebraic concepts was analyzed for possible learning-units, of which the author located a total of 770 in the following order: addition, 125; subtraction, 130; multiplication, 154; division, 77; parentheses, 19; algebraic concepts, 36; and others in combination of addition and subtraction, and in the use of algebraic shorthand.

On the basis of this analysis, tests were constructed with one or more representative problems for each of the above identified learning-units. These tests were given as soon as the particular part of the text corresponding to the test had been completed by the class. For example, as soon as they had covered the portion of the text relating to addition of known numbers, the corresponding test was given. There were a total of twenty-three tests with coefficients of reliability ranging from .69 to .94, the reliability being determined by correlation of chance halves and the use of Brown's formula.

The high schools cooperating in the test program give a good sampling of first year algebra pupils in the Middle West. They were

Oskaloosa, Albia, and Charles City, Iowa, and Marshall, Missouri—Charles City completing the test program when Marshall dropped out. The textbooks represented were: Wells and Hart, Modern First Year Algebra; Newell and Harper, A Year of Algebra; Schorling and Clark, Modern Algebra, Ninth School Year.

These cities are all small enough to insure a good sampling of rural children. Two are in the coal mining section of Iowa, two are railroad division points, two have small colleges within their environs, and all of them are in a small way manufacturing cities. This should give a good sample of pupil ability as the research includes test data from all pupils enrolled in Algebra I, with the exception of "repeaters." The accompanying table gives the results in terms of raw data on all problems in each test; also with data corrected to wrongs per one hundred pupils.

The data would seem to be worthy of consideration, inasmuch as they are drawn from a minute covering of the field over which the tests extended, every learning-unit being presented as a test problem—in fact, it is probable that no such complete testing program has ever been undertaken in the field of Algebra research. The number of pupils whose test results go to make up the data is sufficiently large to give an adequate sampling of pupil ability.

It will be noted from the table that in only one case did the boys show as few wrong problems per one hundred cases as did the girls, that of horizontal addition of literal numbers, where the boys made 444 wrongs to the girls 480 wrongs. In every other test the girls had the advantage of the boys in terms of correct solving of the problems. The corrected data totals show that the boys made 2096 more errors per hundred pupils than did the girls. Or, putting the results in terms of the average pupil, the average boy made 163.77 errors, to 142.81 errors per average girl.

Inasmuch as Iowa has no compulsory high school attendance law, and the opportunities for slightly over-age boys to work in the mines are quite plentiful, the real differences are probably even greater than the data show, for the girls represent a less highly selected group than the boys. While the author would not infer that the findings should be taken at face value, they do seem to indicate a slight superiority in algebraic ability in favor of the girls.

NUMBER OF WRONG PROBLEMS IN TWENTY-THREE TESTS TAKEN BY BOYS AND GIRLS

Field of test	Boys	Wrongs	Corrected	Girls	Wrongs	Corrected
Vertical addition, knowns	163	593	364	192	688	358
Horizontal addition, knowns	156	528	340	196	489	250
Vertical addition, literals	155	1,031	665	188	1,246	663
Horizontal addition, literals	173	768	444	178	855	480
Addition, polynomials	161	603	375	183	431	236
Vertical subtraction, knowns	176	524	298	181	485	268
Horizontal subtraction, knowns	176	1,467	834	181	1,417	783
Vertical subtraction, literals	169	1,015	601	181	1,048	579
Horizontal subtraction, literals	152	2,084	1,371	187	2,108	1,127
Inverted form	158	510	323	186	519	279
Subtraction, polynomials	165	789	484	183	752	411
Vertical multiplication, knowns	165	221	134	185	208	112
Horizontal multiplication, knowns	165	454	275	185	465	261
Multiplication, literals	161	1,580	981	179	1,396	780
Special products	164	1,122	684	181	1,001	553
Monomials times polynomials	158	2,008	1,271	190	2,080	1,095
Division of knowns	156	648	415	182	548	301
Division of literals	150	2,138	1,425	179	2,332	1,303
Removing parentheses	168	1,861	1,112	192	1,803	939
Addition and subtraction, knowns	165	2,072	1,256	184	1,934	1,051
Addition and subtraction, literals	158	1,781	1,127	187	2,065	1,104
Algebraic shorthand	148	631	426	187	623	339
Algebraic concepts	152	1,781	1,172	184	1,874	1,019
Totals	3,714	26,209	16,377	4,251	26,367	14,281
Average	161+			184+		
Average, wrong problems per boy			163.77	per girl		142.81

Reading first line: 163 boys had a total of 593 wrongs in the test in vertical addition of known numbers. Corrected to number of wrongs per one hundred cases, we have 364 wrongs.

One hundred ninety-two girls had a total of 688 wrongs in the same test. Corrected to wrongs per one hundred cases we find 358. Thus the columns headed "corrected" are comparable—the number of wrong problems per one hundred boys with the number of wrong problems per one hundred girls.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

Problem Tendencies in Children, by Willard C. Olsen. Minneapolis: The University of Minnesota Press, 1930. Pp. 92.

Any work that deals in a reasonably objective and as far as possible quantitative manner with behavior problems is welcome. We do not expect too much of these books individually. We ask only that they contribute something toward techniques that may be increasingly scientific.

Dr. Olsen has used a scale previously published, called here "Schedule A—Behavior Problems Record." This is merely a convenient form on which teachers may record overt behaviors. The frequencies of record are then weighted, and a summation gives the child's general tendency to behave atypically. In addition, in this study "Schedule B—Behavior Rating Scale" has been developed. This is a five point "descriptive phrase" scale for rating thirty-five tendencies. The ratings are weighted on the basis of analysis of numerous ratings. Deviates in either direction from the norm are weighted positively. The composite score gives a general measure of the rater's opinion of the child's tendency to deviation from the norm in behavior. The reliabilities of Schedule B (used with 1537 children in Minneapolis first grades, six hundred twenty-six children in several schools from Grades I–VIII, and some nursery school children) are as follows: Repeated ratings for the scale as a whole the self "r" averages .86; splitting the scale and raising the split half correlation gives .92, correlations between different raters are in the neighborhood of .60.

Schedules A and B measured the same thing to the extent represented by a correlation coefficient of .62 (.73 when corrected for attenuation). Certain clinical cases showed very strikingly high ratings.

Dr. Olsen is sufficiently restrained in his conclusions. He finds associations between scores on the Behavior Rating Scale and (*a*)

sex, (*b*) school achievement, (*c*) intelligence. (Generally boys had higher mean scores than girls, and low intelligence [in these groups] tended to be associated with a higher behavior deviation score.) He finds higher correlations between behavior rating tendencies and achievement test scores than between behavior ratings and intelligence test scores.

No attempt is made to offer completely generalized interpretations of the findings. It may fairly be said that the schedules are interesting, and are fairly reliable and valid as these techniques go. They should certainly be known, used, and experimented with by researchers in this important field. It cannot be held against them that they do not offer much immediately to classroom teachers.

DONALD SNEDDEN.

New York University.

A Point Scale of Performance Tests, Volume I, *Clinical Manual*, by Grace Arthur. New York: The Commonwealth Fund, Division of Publications, 1930. Pp. IX + 82.

This work gives a brief history of the work, a description of the tests, explicit directions for administering them, and tables of norms. Volume II, not yet published, will contain "a full account of the work of standardization together with alternative forms of scoring and their results."

This point scale, as described in Volume I, has two advantages over existing performance composites:

1. It is in two forms. Form I and Form II are standardized to be given one year apart. This, of course, permits of rather satisfactory follow-up retesting.

2. All of the standardization was done on the same group of children. Eleven hundred were used with Form I, and five hundred thirty-five of the same children with Form II one year later. This gives authority to the norms for the two forms of the scale.

The tests making up the scale are, in the main, the familiar ones that are used in the Pintner-Paterson Performance Scale. In addition the Porteus Mazes, the Kohs Block Design Tests, and the Healy Pictorial Completion II have been incorporated in the scale. These various tests were selected on the basis of "their ability to discriminate between successive chronological age levels."

The chief disadvantage of the scale as a whole is its inelasticity. Good age norms are available for the total point score, but not for the separate tests. This means that in case a test is "spoiled" an adjustment (that is not simple) has to be made to the incomplete total score.

It seems to the reviewer that the advantages far outweigh the disadvantages and that the scale should have and will have a very wide clinical usefulness.

DONALD SNEDDEN.
New York University.

The Growing Boy, by Paul Hanly Furfey. New York: The Macmillan Co., 1930. Pp. VIII + 192. $2.00.

Growing up continues to be a popular subject with sociologists and psychologists. Dr. Furfey describes in concise and readable form some of the facts about the physical, mental, and personality traits of one hundred sixty-eight boys at ages six, eight, ten, twelve, fourteen, and sixteen. His description includes the central tendencies of behavior at these ages as well as brief studies of individual boys. These studies reveal the wide variation from the average in the development of individuals of the same life age, and the large amount of overlapping in the development of individuals from age to age. Development during the first six years of life is summarized briefly, and adolescence receives special emphasis. Results of objective studies and opinions of leading authorities are included in the discussion of each level. The author concludes that the age of six marks the transition point from home to school and the beginning of social group contacts. The eight-year-old period is characterized by individualism, the ten-year period marks the beginning of the gang age, at twelve the gang age reaches its height and the boy is in a transition period between childhood and adolescence, the fourteen year level continues to be a period of transition characterized by lack of homogeneity in members of the group studied and by difficulty in adjustment. By the age of sixteen the boy is mentally and physically mature but not yet well-adjusted emotionally.

Some of the author's statements concerning characteristics of stages of growth appear to be more largely a matter of opinion than of scientific fact. Whether the eight-year period is more predominantly individualistic than other periods and what the chronological limits of the gang age really are depend somewhat upon definition and inter-

pretation. More emphasis might well have been given to the continuity of growth and the imperceptible shading from one period to another rather than to growth in terms of differentiated stages. A description of how children learn, how concepts and meanings are acquired from birth to maturity, is almost entirely omitted from the present discussion, though it assuredly deserves a place in any study of growing children. Adequate and well-chosen bibliographies appear at the close of each chapter.

GERTRUDE HILDRETH.

Lincoln School of Teachers College

INCHES
METRIC
C2399